

Predictive Modeling of Student Enrollment in Nepalese Universities: A Case Study Using Statistical and Mathematical Methods

Prem Kumar Nayak ¹, Suresh Kumar Sahani ^{2*}, Shyam Narayan Labh ³, Avishek Sah ^{4*}

¹ Faculty of Science, Technology, and Engineering, Rajarshi Janak University, Janakpurdham, Nepal
premnayak6511@gmail.com

² Faculty of science, technology, and engineering, Rajarshi Janak university, Janakpurdham, Nepal
sureshsahani@rju.edu.np

³ President, Nepal Aquaculture Society
snlabh@gmail.com

⁴ Faculty of science, technology, and engineering, Rajarshi Janak university, Janakpurdham, Nepal
avisheksah@rju.edu.np

ARTICLE INFO

Keywords:

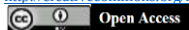
student enrollment; predictive modeling; Nepal; higher education; regression; forecasting; time series; UGC; statistical analysis; mathematical modeling.

Vol. 01, No. 05, 2026

<https://doi.org/10.5281/zenodo.23275665>

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

Student enrollment is one of the most visible indicators of the expansion and changing demand for higher education. However, counts reported for a single year are not sufficient for planning classrooms, faculty, scholarships, digital services, examinations, hostels, and public financing. This study develops a transparent predictive framework for university enrollment in Nepal by combining descriptive statistics, concentration measures, trend analysis, linear regression, exponential regression, growth rates, and scenario projection. The empirical case study uses official University Grants Commission (UGC) data, with the 2023/24 system-level total of 633,053 students and university-level enrollment observations for selected institutions. The seven focal universities are Tribhuvan University (TU), Purbanchal University (PU), Pokhara University (PoKU), Kathmandu University (KU), Far Western University (FWU), Mid-West University (MU), and Rajarshi Janak University (RJU).

The descriptive results show an exceptionally concentrated enrollment structure: TU accounts for 491,299 students, or 77.61% of the national total reported in the UGC dataset, while the seven selected universities together account for approximately 97.68%. Female students constitute 56.57% of the national enrollment in the reported data, although university-level gender patterns differ. For predictive modeling, the paper uses a longer annual series available for TU and KU from 2015/16 through 2023/24, with the 2022/23 observation unavailable in the selected UGC table. Linear and exponential models are estimated and compared using R^2 and RMSE, while model parsimony and policy plausibility are considered before selecting a baseline forecasting approach. The resulting forecasts are not treated as certainties; they are planning estimates conditional on historical patterns continuing.

The study finds that simple linear trend models provide a useful conservative baseline, whereas exponential models generate higher forecasts when sustained proportional growth is assumed. For TU, the forecasts diverge substantially, demonstrating why scenario ranges are more informative than a single point estimate. For KU, the models produce much closer forecasts, reflecting a more stable historical trajectory. At the national level, a two-observation growth scenario based on 2021/22 and 2023/24 implies an average annual growth rate of about 4.52%; because only two national observations are used, this is explicitly treated as a scenario rather than a statistically validated forecast. The paper recommends institutional-level longitudinal data, cohort tracking, program-level predictors, and rolling forecast evaluation as the next stage of Nepalese higher-education planning.

Introduction

Higher education systems are dynamic populations. Students enter through multiple pathways, progress at different speeds, change programmes, pause their studies, graduate, transfer, or leave without completing. Consequently, the number of students enrolled in a given academic year is the result of several interacting flows rather than a single decision. For universities, understanding these flows is

important because enrollment determines a substantial part of the demand for teaching, assessment, advising, information systems, libraries, laboratories, and student support. For governments and regulatory agencies, enrollment forecasts support medium-term resource allocation and infrastructure planning.

Nepal has experienced substantial institutional expansion and diversification. The UGC Higher Education EMIS Report 2023/24 reports 633,053 students across 1,432 campuses in its higher-education reporting structure. The same source shows that TU alone has 491,299 students, illustrating the unusually high concentration of students in one institution. The source also reports 274,913 male and 358,140 female students nationally, giving females a majority of the reported enrollment. These figures provide an appropriate starting point for a mathematical study because they contain both aggregate and institution-level dimensions. The source document supplied for this study likewise emphasizes enrollment concentration, gender composition, geographic distribution, and the need for multidimensional analysis. □ filecite□turn0file0□L52-L63□

The central purpose of this paper is not simply to extrapolate a line into the future. A useful predictive model must answer three questions: what pattern is present in the historical data, how well does a mathematical model reproduce that pattern, and under what assumptions should its forecast be used? This distinction is particularly important in Nepal, where enrollment can be influenced by changes in examination schedules, new universities and programs, international study decisions, demographic changes, migration, tuition costs, labor-market expectations, and policy interventions.

Problem Statement

Despite the importance of enrollment forecasting, institutional planning can be dominated by descriptive counts and short-term comparisons. A university may observe a large increase in enrollment without knowing whether the increase is persistent, temporary, or concentrated in a particular program. Likewise, a decline may reflect a structural fall in demand or a temporary administrative disruption. Without a longitudinal modeling framework, these possibilities are difficult to distinguish.

Research Gap

Existing Nepalese higher-education discussions frequently describe enrollment levels, access, student performance, and institutional challenges separately. The present study adds a predictive component by connecting official enrollment data to explicit mathematical models and by comparing alternative forecasting assumptions. The contribution is methodological: the study demonstrates a reproducible workflow rather than claiming that one model can fully explain national student behavior.

Objectives, Questions and Hypotheses

General Objective

To develop and evaluate a transparent statistical and mathematical framework for predicting student enrollment in Nepalese universities, using official higher-education data and selected universities as a case study.

Specific Objectives

- To describe the current enrollment structure of selected Nepalese universities using official 2023/24 data.
- To quantify institutional concentration through enrollment shares and cumulative concentration ratios.
- To examine male–female enrollment composition and calculate gender participation indicators.

- To model longitudinal enrollment trends for selected universities with linear and exponential regression.
- To compare models using goodness-of-fit and error measures, while considering parsimony and plausibility.
- To generate short- to medium-term enrollment forecasts and alternative scenarios.
- To identify data requirements and policy implications for improved university enrollment planning.

Research Questions

- What is the present distribution of students among selected Nepalese universities?
- How concentrated is enrollment in the largest institution?
- Do historical enrollment patterns support a stable trend model for forecasting?
- How different are forecasts produced by linear and exponential models?
- What are the practical limitations of predicting enrollment from historical counts alone?
- How can universities use forecasts without treating them as certain outcomes?

Working Hypotheses

H1: University enrollment in Nepal exhibits statistically meaningful temporal trends that can be represented by simple regression models.

H2: Alternative functional forms produce materially different forecasts when enrollment growth is rapid or irregular.

H3: A scenario-based forecast range is more appropriate for planning than a single deterministic point forecast when the historical series is short.

Scope

The study focuses on student enrollment rather than academic quality or causal determinants. It uses official secondary data and does not claim to estimate individual-level behavior. The longitudinal predictive analysis is concentrated on TU and KU because the selected UGC report provides sufficiently long comparable series for these institutions. Cross-sectional 2023/24 comparisons are presented for seven universities.

Review of Literature and Conceptual Background

Enrollment forecasting is part of a broader literature on educational planning, student persistence, demographic demand, and time-series prediction. Human-capital theory treats education as an investment in productive capacity, while student-persistence theories emphasize that continuation depends on academic, social, institutional, and personal processes. Tinto's framework, for example, links persistence to academic and social integration, while Bean's attrition model emphasizes a wider set of organizational and behavioral influences. These theories matter for predictive modeling because enrollment is not an isolated variable; it is an observable outcome of multiple underlying processes.

Statistical forecasting literature provides a complementary perspective. Regression models are useful when the researcher wants to represent a systematic relationship between an outcome and time or explanatory variables. A linear trend assumes that enrollment changes by an approximately constant number of students per unit of time. An exponential trend instead assumes that enrollment changes approximately in proportion to its current level. The latter is appropriate only when proportional growth is substantively plausible. Time-series forecasting texts also emphasize the need to examine trend, seasonality, autocorrelation, structural breaks, and forecast uncertainty rather than relying on fit alone.

For Nepalese higher education, the UGC EMIS reports are particularly important because they provide standardized information on institutions, enrollment, gender, teachers, graduates, and other system characteristics. The UGC report used in the present study records both the scale of the national system and substantial differences between institutions.

Recent studies have highlighted the importance of mathematical modelling, numerical methods, and computational approaches in addressing complex problems across different disciplines. Sah et al. (2024) examined the education system of Nepal and identified major challenges requiring systematic and analytical solutions. Similarly, Sahani et al. (2024) demonstrated the application of differential equation frameworks for analysing complex engineering systems, emphasizing the role of mathematical approaches in understanding dynamic processes.

Further studies by Sahani and colleagues focused on the application of numerical methods in social and cultural systems. Their research explored cultural value conflicts, optimization of cultural education programs, and the influence of education policies on literacy improvement through numerical modelling approaches (Sahani & Karna, 2024, 2025; Sahani & Prasad, 2023; Sahani & Sah, 2025; Sahani et al., 2025). These studies indicate that computational techniques can be effectively applied beyond traditional engineering fields to analyse human, educational, and societal phenomena.

Recent research has expanded these approaches by integrating machine learning with analytical and numerical frameworks. Sahani et al. (2026) proposed hybrid models for risk management, uncertainty quantification, credit risk analysis, and intelligent decision-making in Nepalese management systems. Their studies emphasized the potential of explainable machine learning, combined with mathematical modelling, to improve prediction accuracy and support strategic decisions. Similar hybrid architectures have also been applied in healthcare management, dental practice systems, and resource allocation, demonstrating the broader applicability of computational intelligence in complex real-world environments (Sahani et al., 2026).

Overall, the reviewed literature shows a gradual shift from traditional analytical and numerical methods toward integrated hybrid frameworks combining mathematics, simulation, and machine learning. These approaches provide improved capability for modelling uncertainty, optimizing decisions, and solving multidisciplinary problems. However, further research is needed to validate these models using larger datasets and real-world applications.

Conceptual Model

The conceptual model used here treats observed enrollment E_t as the result of a historical trend plus unobserved shocks. In a simple deterministic trend model, $E_t = a + bt + \varepsilon_t$, where a is the intercept, b is the average change per year, and ε_t represents deviations from the trend. In an exponential model, $\ln(E_t) = \alpha + \beta t + \varepsilon_t$, so $E_t = \exp(\alpha + \beta t)$. The parameter β can be interpreted as an approximate proportional growth rate under the model.

Why Multiple Models?

No single functional form should be assumed in advance. A linear model can be robust for planning when institutional capacity places practical limits on growth. An exponential model may better describe a period of sustained proportional expansion but can become unrealistic over a long horizon. Therefore, the study compares both models and reports the divergence between their forecasts.

Data and Methodology

Research Design

The study adopts a quantitative secondary-data design. The main source is the University Grants Commission Nepal Higher Education EMIS Report 2023/24, supplemented by UGC historical reporting and established statistical and forecasting literature. The design has two components. First, a cross-sectional component compares seven universities in 2023/24. Second, a longitudinal component models enrollment over time for TU and KU using annual observations reported in the UGC series.

Data Structure

Variable	Definition	Use
E _t	Total enrollment in year t	Outcome variable
Male _t	Male students in year t	Gender comparison
Female _t	Female students in year t	Gender comparison
S _i	University i share of national enrollment	Concentration
t	Year index	Trend predictor
g	Annual/CAGR growth rate	Scenario projection

Study Universities

The selected institutions represent different scales and institutional contexts: TU, PU, PoKU, KU, FWU, MU and RJU. The selection is purposive and should not be interpreted as a statistically representative sample of all universities. It is designed to show the range from a very large national system to an emerging university. The source study similarly used these seven institutions for comparative mathematical analysis. □ filecite□turn0file0□L183-L195□

Data Quality Considerations

Historical educational datasets may contain revisions, changes in reporting coverage, changes in affiliation structure, and missing observations. In the selected UGC table, the longitudinal series lists 2015/16 through 2021/22 and 2023/24 for TU and KU, leaving 2022/23 absent in that table. The analysis therefore does not invent a 2022/23 value. The missing year is retained as a gap in the time axis. This is important because interpolating an administrative observation without documentation could create false precision.

Ethical and Reproducibility Statement

The analysis uses aggregated institutional data and does not involve identifiable student records. All transformations are formula-based and can be reproduced from the reported enrollment counts. Forecasts are labeled estimates or scenarios rather than official projections.

Mathematical and Statistical Methods

Enrollment Share

For university i, enrollment share is $S_i = (E_i / E) \times 100$, where E_i is enrollment of university i and E is national enrollment.

For TU: $S_{TU} = (491,299 / 633,053) \times 100 = 77.61\%$.

Gender Participation and Gender Parity

Female participation is $F_i = \text{Female}_i / E_i \times 100$. A gender parity ratio is $GPI_i = \text{Female}_i / \text{Male}_i$. Values above 1 indicate a numerical female majority.

Growth Rate and CAGR

The annual growth rate is $g_t = (E_t - E_{t-1}) / E_{t-1}$. For a multi-year interval, $CAGR = (E_t / E_0)^{(1/n)} - 1$.

Linear Trend Model

The linear model is $E_t = a + bt + \varepsilon_t$. The least-squares estimates minimize $\sum (E_t - \hat{E}_t)^2$. The slope b represents the estimated average annual change in enrollment.

Exponential Trend Model

The exponential model is $E_t = A \exp(Bt) + \varepsilon_t$ in multiplicative form, or $\ln(E_t) = \alpha + \beta t + \varepsilon_t$ after logarithmic transformation. The implied proportional growth rate is approximately $\exp(\beta) - 1$.

Goodness-of-Fit Measures

$R^2 = 1 - SSE/SST$ measures the proportion of variation reproduced by the fitted model. $RMSE = \sqrt{(1/n) \sum (E_t - \hat{E}_t)^2}$ measures the typical magnitude of prediction error in students. R^2 alone is not sufficient for model selection because a more flexible model can fit historical observations well while generating implausible forecasts.

Forecast Scenarios

Forecasts are produced for 2024/25–2028/29. A baseline linear forecast is reported alongside an exponential scenario. The two models define a planning range rather than a confidence interval. A true statistical prediction interval would require a specified stochastic time-series model and a formal uncertainty calculation; the present study avoids overstating precision.

National Scenario

Because the available national observations used here are 579,448 students in 2021/22 and 633,053 in 2023/24, a two-point CAGR of approximately 4.52% is calculated only as an illustrative scenario. It is not treated as a validated national forecast.

Descriptive Results: 2023/24 Enrollment

The 2023/24 data show a highly uneven distribution of students across the seven selected universities. TU is far larger than every other institution, while RJU is still at an emerging scale. This pattern has direct implications for forecasting: a national total can be strongly affected by a single institution, while small universities can show large percentage changes from relatively small absolute changes.

University	Enrollment	National share	Male	Female	Female share
TU	491,299	77.61%	204,449	286,850	58.39%
PU	39,780	6.28%	19,102	20,678	51.98%
PoKU	35,200	5.56%	19,560	15,640	44.43%
KU	21,053	3.33%	11,148	9,905	47.05%
FWU	19,356	3.06%	6,530	12,826	66.26%
MU	11,171	1.76%	4,736	6,435	57.60%
RJU	527	0.08%	240	287	54.46%

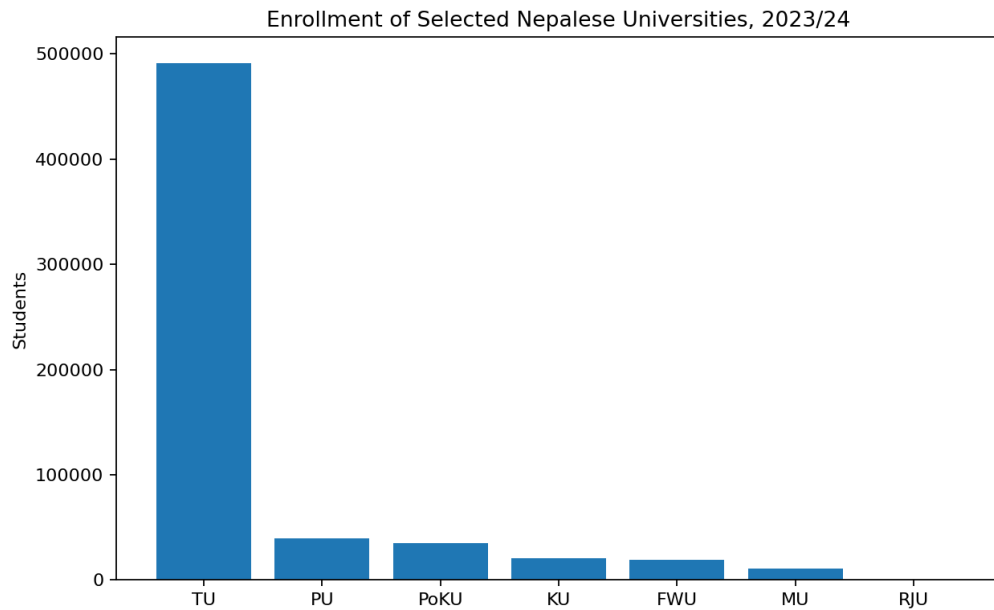


Figure 1. Enrollment of selected Nepalese universities, 2023/24.

The table reproduces the principal enrollment values in the supplied source, including the national shares and gender composition. The seven universities collectively account for approximately 97.68% of the national enrollment in the comparison dataset. This is a descriptive concentration result, not evidence that the remaining institutions are unimportant; rather, it shows that the selected group contains most of the reported student volume.

National Gender Pattern

The national totals reported by UGC are 274,913 male and 358,140 female students. The female share is therefore 56.57%, while the male share is 43.43%. However, the university-level figures show heterogeneity: female participation is highest in FWU and lowest in PoKU among the selected institutions.

Concentration and Gender Analysis

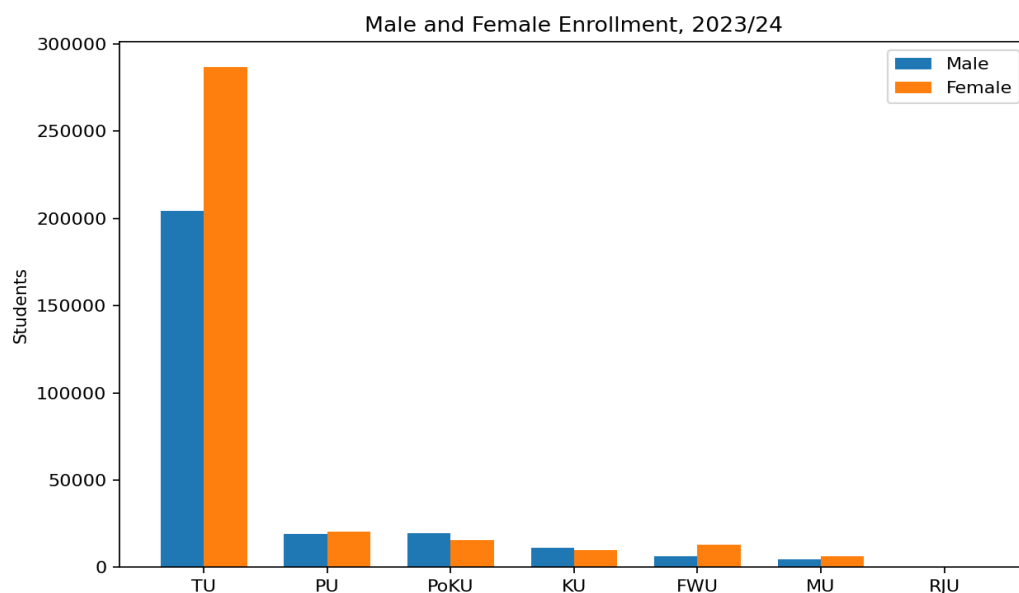


Figure 2. Male and female enrollment in the selected universities.

Concentration Ratios

CR1 = 77.61% for TU. CR3 = 77.61 + 6.28 + 5.56 = 89.45%. CR4 = 92.78%. Thus, the three largest selected universities contain nearly nine-tenths of the national enrollment reported in the dataset.

Measure	Formula	Result
CR1	TU / National	77.61%
CR3	TU + PU + PoKU	89.45%
CR4	TU + PU + PoKU + KU	92.78%
CR7	Seven selected / National	97.68%

Herfindahl-Type Concentration

A concentration measure over the seven selected institutions is $HHI_7 = \sum s_i^2$. Using the reported percentage shares gives approximately 6,116.9 percentage-point-squared. Because the calculation excludes other universities, it should not be interpreted as a conventional market-competition HHI; its purpose is to quantify dominance by TU within the selected set.

Gender Parity

University	Female/Male ratio	Interpretation
TU	1.40	Female majority
PU	1.08	Slight female majority
PoKU	0.80	Male majority
KU	0.89	Male majority
FWU	1.96	Strong female majority
MU	1.36	Female majority
RJU	1.20	Female majority

The gender results confirm that a national female majority should not be interpreted as uniform parity across institutions. The supplied source reaches the same conclusion and recommends moving beyond enrollment parity toward programme, completion, employment and leadership outcomes.

□filecite□turn0file0□L337-L365□

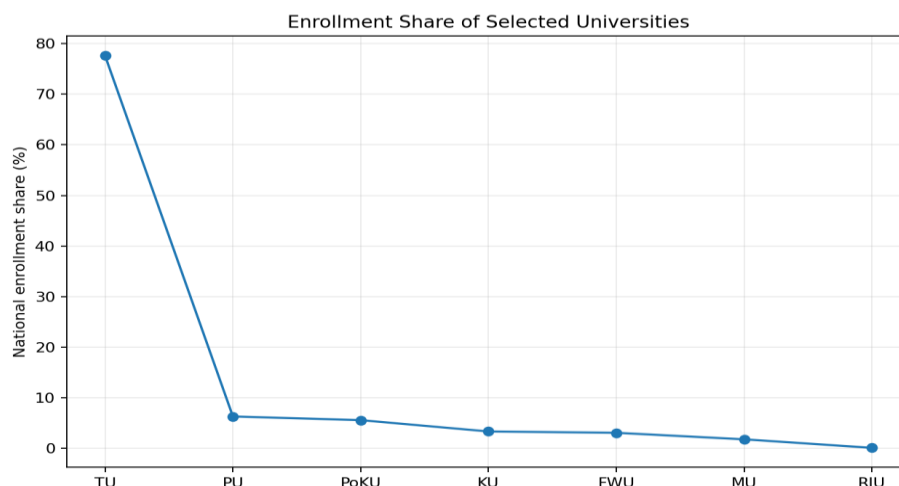


Figure 3. National enrollment shares of selected universities.

Historical Enrollment Patterns for Predictive Modeling

For predictive modeling, the UGC historical table provides a particularly useful series for TU and KU. TU enrollment rises from 284,453 in 2015/16 to 491,299 in 2023/24, with noticeable fluctuations and a marked increase by 2021/22. KU increases more gradually from 16,658 to 21,053 over the same

listed observations. The contrast provides a useful test of model behavior: a highly dynamic series and a relatively stable series should not be expected to produce identical forecasting performance.

Academic year	TU	KU
2015/16	284,453	16,658
2016/17	289,566	17,080
2017/18	346,606	17,942
2018/19	335,543	18,356
2019/20	356,654	18,643
2020/21	347,269	16,046
2021/22	454,065	17,945
2022/23	—	—
2023/24	491,299	21,053

The official UGC report itself presents these historical series. For TU, the reported enrollment rises sharply to 454,065 in 2021/22 and 491,299 in 2023/24. For KU, the series is much smoother overall, although it also contains a decline in 2020/21. [cite](#) [turn0search44](#)

Interpretation of the Missing Year

The absence of 2022/23 in the selected historical table is methodologically important. A forecasting model should not silently fill missing administrative data. The regression uses the actual observed year index, so the interval from 2021/22 to 2023/24 spans two academic-year steps. This prevents the model from treating the gap as if it were an observed annual value.

Trend Characteristics

TU exhibits a larger absolute trend and substantially greater variance than KU. KU's observations are comparatively close to a smooth upward path. Consequently, model selection for TU must give greater attention to structural changes and the danger of extrapolating a recent jump indefinitely.

Regression Model Estimation and Evaluation

Linear Model Results

Institution	Linear R^2	Linear (students)	RMSE	Interpretation
TU	0.866	25,095		Strong upward trend with substantial deviations
KU	0.395	1,111		Moderate trend; comparatively stable scale

For TU, the linear model explains about 86.6% of the observed variation and has an RMSE of approximately 25,095 students. For KU, R^2 is about 0.395 and RMSE is approximately 1,111 students. The lower KU R^2 does not mean that the model is useless: the absolute forecast error is much smaller because KU's enrollment scale is much smaller and the series contains less pronounced growth.

Exponential Model Results

Institution	Exponential R^2	Exponential (students)	RMSE	Implied growth tendency
TU	0.886	23,115		Proportional growth

KU	0.402	1,105	Mild proportional growth
----	-------	-------	--------------------------

The exponential model fits TU with $R^2 \approx 0.886$ and $RMSE \approx 23,115$ students, slightly improving historical fit relative to the linear model. For KU, the exponential model has $R^2 \approx 0.402$ and $RMSE \approx 1,105$, only marginally different from the linear model. The modest improvement for TU should not automatically justify exponential extrapolation because the forecast can accelerate rapidly.

Model Selection Principle

The study adopts the linear model as the conservative baseline because it is simple, interpretable, and less likely to produce explosive long-range forecasts. The exponential model is retained as an upper-growth scenario. This decision reflects both statistical fit and substantive plausibility rather than R^2 alone.

The modeling results should be interpreted as conditional estimates. They do not establish causal relationships between enrollment and demographic, economic, policy, or migration variables.

Forecast Results: Tribhuvan University

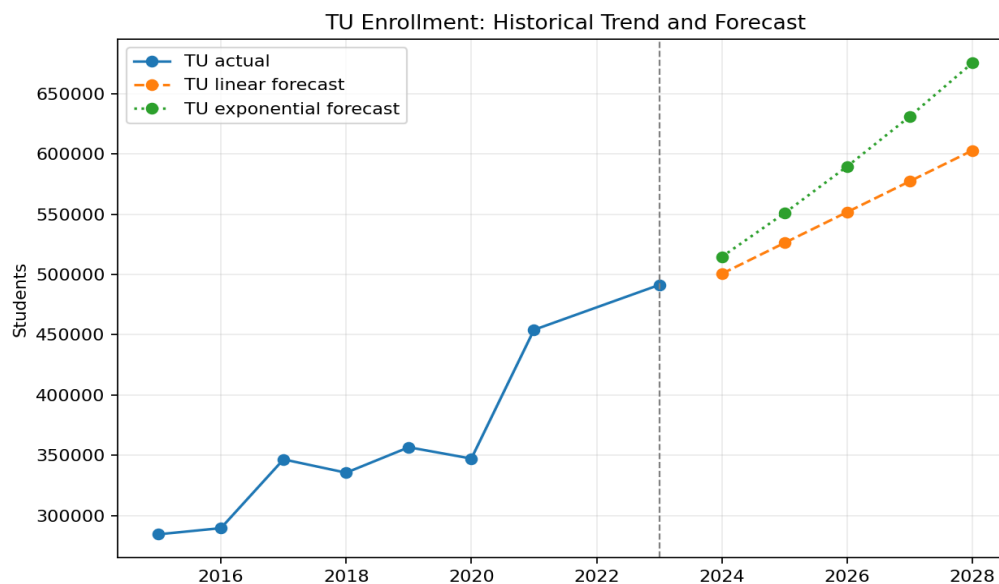


Figure 4. TU historical enrollment and 2024–2028 forecasts under linear and exponential models.

Academic year	Linear baseline	Exponential scenario	Difference
2024/25	500,574	514,621	14,046
2025/26	526,136	550,812	24,676
2026/27	551,697	589,548	37,850
2027/28	577,259	631,008	53,749
2028/29	602,820	675,384	72,563

The TU forecasts demonstrate the central practical lesson of this study. The linear baseline rises from about 500,574 students in 2024/25 to 602,820 in 2028/29. The exponential scenario rises faster, from about 514,621 to 675,384. The difference between the two scenarios reaches more than 72,000 students by the end of the horizon. This is not an error; it is a direct consequence of different assumptions about growth. The linear model assumes a roughly constant annual increase, whereas the exponential model assumes a roughly constant proportional growth rate.

For planning, the baseline should not be interpreted as a guaranteed ceiling. Instead, it provides a conservative reference against which capacity requirements can be tested. If admissions, new programs,

demographic demand, or administrative changes cause proportional growth to persist, the exponential scenario indicates that capacity requirements could be substantially higher. The prudent policy response is therefore to monitor actual enrollment annually and update the model rather than commit infrastructure solely to one five-year estimate.

Forecasting Equation

Using the fitted linear relationship $E_t = a + bt$, the predicted value is generated by inserting the future year index. The corresponding exponential scenario uses $\ln(E_t) = \alpha + \beta t$ and back-transforms through $E_t = \exp(\alpha + \beta t)$.

Forecast Results: Kathmandu University

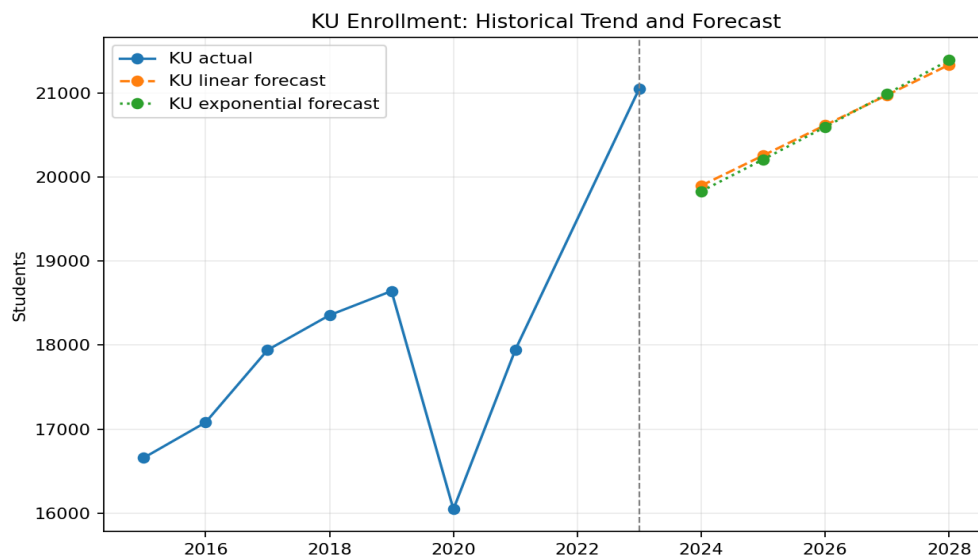


Figure 5. KU historical enrollment and 2024–2028 forecasts under linear and exponential models.

Academic year	Linear baseline	Exponential scenario	Difference
2024/25	19,898	19,831	-67
2025/26	20,257	20,210	-47
2026/27	20,617	20,597	-20
2027/28	20,977	20,990	14
2028/29	21,336	21,392	56

KU produces a very different forecasting picture. The linear baseline increases from approximately 19,898 students in 2024/25 to 21,336 in 2028/29. The exponential scenario produces approximately 19,831 to 21,392 over the same period. The gap remains small because the underlying historical trajectory is comparatively stable. This convergence increases confidence in the direction of the forecast, although it does not remove uncertainty caused by future changes in programs, admissions, migration, fees, or institutional strategy.

Comparing TU and KU

The two cases show why a national forecasting model should not automatically be applied identically to every institution. TU's enrollment is nearly 932 times RJU's enrollment in 2023/24, according to the source data, and its historical series contains a much larger absolute movement. KU operates at a smaller scale and has a more gradual trajectory. A single growth rule can therefore be misleading when institutions differ greatly in mission, affiliation structure, and scale.

□ filecite □ turn0file0 □ L352-L355 □

Planning Interpretation

For a stable institution like KU, the difference between linear and exponential scenarios is relatively limited over five years, so capacity planning can use a narrow scenario band. For TU, the gap is large enough to justify a rolling forecasting process: update the model each year, test recent residuals, and use alternative scenarios for admissions, faculty, classrooms and examination capacity.

The results also demonstrate that statistical significance is not the same as practical usefulness. A model may fit historical data well but still be inappropriate for a structural change. Forecasts should therefore be paired with expert review and institutional information.

National Enrollment Scenario Projection

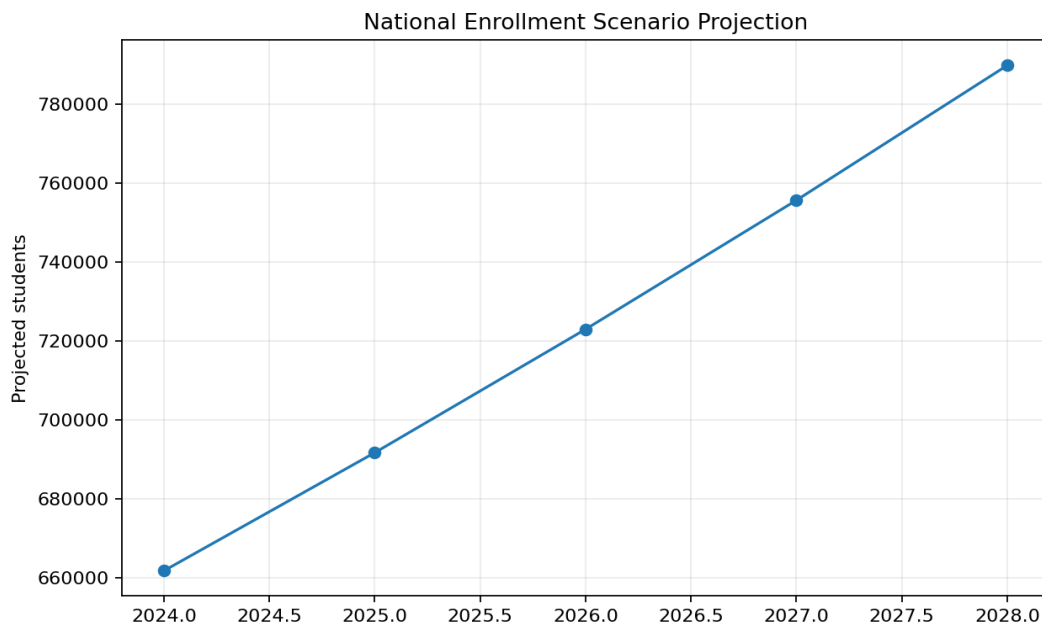


Figure 6. Illustrative national enrollment scenario based on the 2021/22–2023/24 CAGR.

The UGC historical material reports 579,448 higher-education students for 2021/22 in one national series and the 2023/24 report reports 633,053 students. Using these two observations, the implied two-year CAGR is: $CAGR = (633,053 / 579,448)^{(1/2)} - 1 \approx 0.04523$, or 4.52% per year. Under a constant-growth scenario, the projected totals are approximately 661,687 in 2024/25, 691,617 in 2025/26, 722,900 in 2026/27, 755,599 in 2027/28 and 789,776 in 2028/29.

Year	Scenario enrollment
2023/24 (base)	633,053
2024/25	661,687
2025/26	691,617
2026/27	722,900
2027/28	755,599
2028/29	789,776

Why This Is a Scenario, Not a Forecast

Only two national observations are used in this calculation. A two-point CAGR cannot test trend stability, seasonality, structural breaks, or model residuals. It also cannot identify the causes of growth. Therefore, the numbers are best used as a planning scenario: if the average annual proportional growth

observed between the two reported points continued unchanged, the system would approach 790,000 students by 2028/29. This should not be presented as an official prediction.

Implications

Even a modest sustained growth rate creates large absolute capacity requirements because the starting population is already large. At 4.52% annual growth, an additional roughly 157,000 students would be implied over five years relative to the 2023/24 base. The actual outcome could be lower or higher. Universities should therefore maintain capacity buffers and update forecasts using the latest UGC data each year.

The UGC strategic planning context also emphasizes the need for stronger higher-education financing and more balanced institutional development. □cite□turn1search15□

Predictive Modeling Framework and Diagnostic Logic

Predictive Modeling Framework for University Enrollment

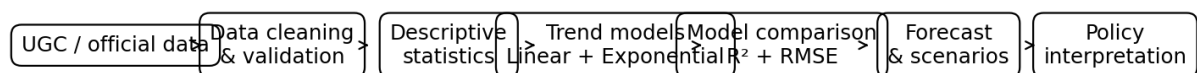


Figure 7. Reproducible predictive modeling workflow.

Step 1: Data Acquisition

Collect comparable annual enrollment observations from UGC EMIS and university records. Record the academic year, reporting definition, institution, program coverage and any revisions.

Step 2: Data Validation

Check duplicates, missing years, implausible jumps, changes in affiliation, and consistency between male, female and total counts. Missing observations should remain missing unless a documented imputation procedure is justified.

Step 3: Exploratory Analysis

Plot enrollment against time, compute annual growth rates, inspect concentration, and compare institutions. This stage can reveal whether a linear, exponential, or more flexible model is plausible.

Step 4: Model Estimation

Estimate candidate models. For a linear trend, fit $E_t = a + bt$. For exponential growth, fit $\ln(E_t) = \alpha + \beta t$. If sufficiently long data become available, additional candidates could include Holt's trend method, ARIMA, dynamic regression, or models with demographic and economic predictors.

Step 5: Evaluation

Use out-of-sample testing whenever the data length permits. Compare MAE, RMSE, MAPE where appropriate, and examine residual patterns. With very short series, cross-validation can be unstable; therefore, simple models may be preferable.

Step 6: Forecast and Update

Generate a baseline forecast and alternative scenarios. Each year, append the new observation and re-estimate the model. A forecast is therefore a living planning instrument rather than a fixed statement about the future.

Step 7: Policy Translation

Translate student forecasts into classrooms, teachers, library seats, laboratory capacity, digital accounts, examination loads and budget scenarios. Forecasting becomes useful only when it changes planning decisions.

Discussion

The results provide three main findings. First, the present higher-education system is highly concentrated. TU's 77.61% national enrollment share means that national aggregates are strongly influenced by one institution. This concentration is historically rooted in TU's large affiliated network, but it creates forecasting and management risks: a small percentage change at TU corresponds to a large absolute change nationally. The source material similarly describes TU as a system-within-the-system and argues that its performance has disproportionate effects on national indicators.

Second, university trajectories are heterogeneous. TU's historical series shows substantial increases and fluctuations, while KU's series is comparatively gradual. The predictive models capture this difference. For TU, linear and exponential forecasts diverge rapidly, meaning that model assumptions become increasingly consequential. For KU, the two models remain close. This suggests that forecasting should be institution-specific or peer-group-specific rather than based on a single national growth coefficient.

Third, prediction should be separated from explanation. The models in this paper use time as the principal predictor. They do not explain why enrollment changes. Possible explanatory variables include secondary-school completion, population in the university-age group, household income, tuition and scholarship policies, employment expectations, migration and foreign-study decisions, program availability, examination schedules, and institutional reputation. Future research should integrate such variables into multivariate or panel models.

Relation to Student Persistence

Enrollment forecasting also needs a cohort perspective. Annual enrollment combines continuing students and new entrants. A rise in total enrollment can occur because more students are entering, because students are taking longer to complete, or because both are happening. Persistence theories therefore complement aggregate forecasting. Tinto and Bean demonstrate why continuation and attrition should be modeled as processes rather than inferred from annual totals alone.

Gender and Equity

The national female majority is an encouraging access indicator, but it should not be interpreted as complete equity. The source data show female majorities in TU, PU, FWU, MU and RJU, but male majorities in PoKU and KU. Programme choice, graduation, employment and leadership may reveal different inequalities. A future predictive model should therefore forecast not only total enrollment but also gender-disaggregated and programme-disaggregated demand.

Policy and Planning Implications

Establish a National Enrollment Forecasting Dashboard

UGC and universities should maintain a common annual forecasting dashboard containing total enrollment, new entrants, continuing students, graduates, withdrawals, transfers, female and male enrollment, program-level demand, and institutional capacity. Each indicator should display its denominator and reporting year. Machine-readable data would enable independent verification and more frequent model updating.

Use Scenario Planning Rather Than One Number

At minimum, planners should maintain conservative, baseline and high-growth scenarios. The TU results demonstrate why this matters: a linear and exponential model can produce a difference of tens of thousands of students within a five-year horizon. Scenario planning reduces the risk of either overbuilding or underpreparing.

Link Enrollment to Capacity

Enrollment forecasts should be translated into student-to-teacher ratios, classroom demand, laboratory sessions, library use, digital learning accounts, examination workload, hostel demand and operating expenditure. A forecast without a capacity response has limited practical value.

Strengthen Regional Universities

Regional universities can reduce geographic concentration and provide locally relevant programs. However, expansion should be matched with faculty, laboratories, libraries, digital infrastructure and quality assurance. The supplied source recommends targeted support for regional and emerging universities rather than simply replicating the scale of TU. [filecite\turn0file0\L481-L488](#)

Improve Student Success Data

Universities should track cohorts from admission through graduation. A completion ratio can be defined as $C = \text{graduates} / \text{relevant entering cohort} \times 100$. Pass rates can be calculated as $\text{students passing} / \text{students appearing} \times 100$. Cohort indicators are more informative than annual counts because they distinguish genuine progression from delayed completion. [filecite\turn0file0\L431-L435](#)

Integrate Predictors

Once a longer panel becomes available, regression can include demographic and institutional predictors. A general model could be written as $E_{it} = \alpha_i + \beta_1 \text{Pop}_{it} + \beta_2 \text{Program}_{it} + \beta_3 \text{Cost}_{it} + \beta_4 \text{Income}_t + \beta_5 \text{Migration}_t + u_{it}$, where i indexes university and t indexes year. Such a model would move the study from trend extrapolation toward explanatory prediction.

Limitations and Future Research

Limitations

- The longitudinal case study is based on a relatively short annual series for TU and KU, with 2022/23 absent from the selected UGC table.
- Simple trend models do not capture causal drivers, seasonality, structural breaks, or policy shocks.
- The national scenario uses only two reported national enrollment points and therefore cannot provide a statistically validated long-term forecast.
- University selection is purposive; results should not be generalized to every Nepalese university without additional evidence.
- Administrative definitions and reporting coverage can change over time, potentially affecting comparability.
- Forecasts are point scenarios. Formal prediction intervals require a specified stochastic model and stronger time-series assumptions.

Future Research Agenda

- Build a 10–15-year panel for all Nepalese universities and major program clusters.
- Use cohort-level admissions, progression, graduation and withdrawal data.
- Compare linear, exponential, Holt, ARIMA and machine-learning models only when data length supports them.

- Introduce demographic, economic, migration, tuition, scholarship and labor-market predictors.
- Evaluate forecasts using rolling-origin out-of-sample validation.
- Develop gender-, region-, and discipline-specific enrollment forecasts.
- Publish reproducible data dictionaries and model code so annual forecasts can be independently audited.

Recommended Statistical Extension

With a sufficiently long panel, a fixed-effects or mixed-effects model could separate university-specific baseline differences from common time effects. Forecast accuracy could be compared using MAE and RMSE on rolling holdout periods. If quarterly or monthly admission data become available, seasonal models may be appropriate. Machine-learning methods such as random forests or gradient boosting should be considered only when there are enough observations and meaningful predictors; otherwise, complex models may overfit.

Conclusion

This study developed a mathematical and statistical framework for predicting student enrollment in Nepalese universities. The analysis began with the current structure of higher education and then moved from descriptive statistics to concentration measures, historical trend estimation, alternative regression forms, and scenario forecasting. The approach demonstrates that predictive modeling can be useful even when data are limited, provided that assumptions and limitations are made explicit.

The 2023/24 UGC data show that Nepal's higher-education enrollment is strongly concentrated. TU accounts for 77.61% of the 633,053 students reported in the national dataset, and the seven selected universities together account for approximately 97.68%. Female students form 56.57% of the national enrollment, but university-level gender patterns vary. These findings establish that enrollment planning cannot be based on a single national average.

The predictive case studies of TU and KU show why model comparison matters. For TU, the linear and exponential models both fit the historical trajectory reasonably well, but their five-year forecasts diverge substantially. The linear model gives a conservative planning baseline, while the exponential model represents a higher-growth scenario. For KU, both models produce much closer forecasts. This difference reflects institutional heterogeneity and supports the use of peer-group or institution-specific forecasting.

The national 4.52% CAGR scenario should be interpreted cautiously because it uses only two national observations. Its purpose is to illustrate the scale of planning consequences if recent proportional growth continued. A robust national forecast requires a longer series and preferably explanatory predictors.

The central conclusion is that Nepalese higher-education planning should move from counting students to forecasting flows and capacity needs. Enrollment forecasts should be updated annually, tested against observed outcomes, and connected to teachers, infrastructure, finance, student success and regional equity. The most valuable predictive system will not be the one with the most complicated algorithm; it will be the one that uses transparent data, appropriate models, realistic scenarios and continuous evaluation.

Final Mathematical Insight

Forecasting can be summarized as: Data → Pattern → Model → Validation → Scenario → Decision → Update. Each stage matters. A mathematically elegant model cannot compensate for weak

data, and a statistically accurate forecast has little value if it is not translated into institutional capacity and student-support decisions.

References

1. Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25, 297–308.
2. Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.
3. Baral, R. P. (2022). The digital divide in online learning: A case study of university students in Nepal. *Prithvi Academic Journal*, 5(1).
4. Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12, 155–187.
5. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
6. Duwal, B., & Khonju, L. (2021). Factors affecting students' academic performance: The case of students at community colleges in Bhaktapur District. *The International Research Journal of Management Science*, 5(1), 23–38.
7. Gaihre, M. N. (2023). A study of the factors affecting students' performance in English. *The Journal of Academic Development*, 8(1), 27–44.
8. Ghimire, R. (2023). Factors affecting academic achievement of B.Sc. students of Tribhuvan University constituent campuses in Province 1, Nepal. *Himalayan Journal of Science and Technology*, 7(1), 26–37.
9. Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
10. Hamal, J. B. (2021). Impact of social media on academic performance of graduate students in Kathmandu. *Interdisciplinary Journal of Management and Social Sciences*, 2(2), 37–49.
11. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
12. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer.
13. Khanal, S. P. (2019). Irregular attendance of university students at class and its relation to their academic achievement. *Tribhuvan University Journal*, 33(1), 115–128.
14. Ministry of Finance, Government of Nepal. (2024). *Economic Survey 2023/24*. Government of Nepal.
15. Mishra, A., Choudhary, S., & Shiwakoti, R. (2025). Teaching style and administration on student motivation: A study from Kathmandu Valley. *International Journal of Atharva*, 3(1).
16. National Planning Commission. (2023). *Sustainable Development Goals: Status and Roadmap/indicator framework for Nepal*. Government of Nepal.
17. Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications* (4th ed.). Springer.
18. Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). University of Chicago Press.
19. UNESCO. (2023). *Global Education Monitoring Report 2023: Technology in education: A tool on whose terms? UNESCO*.
20. University Grants Commission Nepal. (2024). *Report of Higher Education 2023/024 A.D. (2080/081 B.S.)*. University Grants Commission, Sanathimi, Bhaktapur.

21. University Grants Commission Nepal. (2025). Annual Report 2023/2024. University Grants Commission, Nepal.
22. University Grants Commission Nepal. (2024). UGC Strategic Plan 2024–2029. University Grants Commission, Nepal.
23. World Bank. (2015). Nepal: Higher Education Reforms Project. World Bank.
24. World Bank. (2019). World Development Report 2019: The changing nature of work. World Bank.
25. World Bank. (2020). The COVID-19 pandemic: Shocks to education and policy responses. World Bank.
26. Sah, P., Sah, R., Sahani, S. K., & Sahani, K. (2024). Study and analysis of education system in Nepal and its challenges along with its solution. *Mikailalsys Journal of Advanced Engineering International*, 1(1), 33–42. <https://doi.org/10.58578/mjaei.v1i1.2792>
27. Sahani, S., Sah, A. K., Jha, A., & Sahani, K. (2024). Analytical frameworks: Differential equations in aerospace engineering. *ALSYSTECH Journal of Education Technology*, 2(1), 13–30. <https://doi.org/10.58578/alsystech.v2i1.2267>
28. Sahani, S. K., & Karna, S. K. (2024). Simulation of the clash between cultural values in heterogeneous society using numerical methods. *Journal of Posthumanism*, 4(3), 543–559. <https://doi.org/10.63332/joph.v4i3.2686>
29. Sahani, S. K., & Karna, S. K. (2025). Program for cultural education optimization by the application of numerical methods. *Journal of Posthumanism*, 5(1), 1611–1627. <https://doi.org/10.63332/joph.v5i1.2687>
30. Sahani, S. K., & Prasad, K. S. (2023). Modeling the impact of cultural education policies on literacy rates using numerical differential equations. *Journal of Posthumanism*, 3(3), 376–394. <https://doi.org/10.63332/joph.v3i3.2905>
31. Sahani, S. K., & Sah, B. K. (2025). Optimization of cultural education program using numerical techniques. *Journal of Posthumanism*, 5(6), 4265–4280. <https://doi.org/10.63332/joph.v5i6.2600>
32. Sahani, S. K., Sah, M. K., Agarwal, S. K., Pareek, V., Parihar, P., & Singh, V. V. (2025). Numerical simulation of cultural value conflict in multicultural societies. *Journal of Posthumanism*, 5(3), 1815–1831. <https://doi.org/10.63332/joph.v5i3.2651>
33. 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. *Journal of Intelligent Decision Making and Information Science*, 3(3s), 1850–1873. <https://doi.org/10.59543/jidmis.v3.971>
34. 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>
35. 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, Article 1549. <https://doi.org/10.59543/jidmis.v3.1549>

36. 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.
37. 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.
38. 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.