

Prediction of GPA-5 Achievers in Bangladesh SSC Exams Using Machine Learning

Executive Summary

This report presents a comprehensive predictive analysis of Bangladesh's Secondary School Certificate (SSC) examination results from 2001 to 2025. By leveraging historical data, the study develops machine learning models to forecast the number of students achieving GPA-5—the highest grade possible. The primary aim is to assist education policymakers, data analysts, and stakeholders in understanding performance trends and using them to make data-driven decisions.

Five regression models were built—Linear Regression, Ridge, Lasso, Decision Tree, and Random Forest. Among these, the Random Forest Regressor showed superior performance with an R^2 score of 0.995 and the lowest RMSE of 3431. This makes it highly suitable for future predictions. The final model allows for dynamic forecasting by accepting user inputs for Year, Total Examinees, and Pass Rate.

1. Introduction

The SSC exam is a critical academic milestone in Bangladesh. Understanding trends in top scorers can help identify systemic improvements or issues in the education system. This project aims to model and predict the number of GPA-5 achievers based on annual metrics such as pass rate and number of examinees.

2. Data Overview

- **Source:** Government educational board reports and publicly available datasets.
- **Time Span:** 2001 to 2025 (2025 data partially available/predicted).
- **Features Used:**
 - Year (int)
 - Total_Examinees (int)

- Pass_Rate (float)
- GPA_5_Count (int)

Some missing values for "Total_Examinees" were interpolated for early years (e.g., 2001).

3. Exploratory Data Analysis (EDA)

- The number of GPA-5 achievers has seen exponential growth from 76 in 2001 to over 260,000 in recent years.
 - The correlation between GPA_5_Count and Year is **0.92**, indicating a strong upward trend over time.
 - Pass_Rate has improved from ~35% to over 90%, directly influencing the increase in GPA-5 achievers.
-

4. Model Building & Evaluation

The following models were developed:

Model	R ² Score	RMSE	Notes
Linear Regression	0.897	16001	Good baseline
Ridge Regression	0.896	16041	Similar to linear
Lasso Regression	0.897	16001	Sparse coefficients

Decision Tree	0.914	14583	Captures non-linearities
---------------	-------	-------	--------------------------

Random Forest	0.995	3431	Best model, highly stable
---------------	--------------	-------------	---------------------------

Random Forest Regressor was selected as the final model due to its high accuracy and robustness.

5. Predictive System Design

A simple input-driven prediction system was implemented where users can provide:

- Year (e.g., 2027)
- Total Examinees
- Pass Rate (%)

Based on these, the Random Forest model accurately predicts the expected GPA_5_Count.

6. Insights & Recommendations

Insights:

1. There is a consistently strong growth in GPA-5 achievers post-2010.
2. Increased pass rates correlate with curriculum or policy improvements.
3. The surge in examinees and quality results may reflect better educational access and equity.
4. Year and Total Examinees are strong predictors; however, regional factors could improve predictions further.

Recommendations:

- Use this model to simulate policy impacts (e.g., what happens if pass rates drop?)
 - Include school-level and regional data in future versions.
 - Extend this predictive system to integrate with government dashboards.
 - Retrain the model annually with new data to ensure accuracy.
-

7. Limitations

- The dataset is aggregated at the national level.
 - Socioeconomic or curriculum-based factors are not considered.
 - Some early-year data had to be interpolated.
-

8. Disclaimer

This analysis is based solely on publicly available historical data and is intended for educational and strategic planning purposes. The models built are data-driven and may not capture real-world changes due to policy, disaster, or other social factors. Results and interpretations should be validated with domain experts before real-world implementation.

Prepared by: Kumari Shivangi

Date: July 2025