Exploring And Predicting CGPA of Bangladeshi Students

Team 2:

DOTAN KATZ, ORAN KEDEM, NOAM YEHOSHUA, BAR ELHAYANI

Research Question

General Question:

Can we predict student CGPA based on learning behavior, demographics, and academic history?

Research questions:

- RQ1: Can CGPA be predicted from academic history?
- RQ2: Can CGPA be predicted from demographic background?
- RQ3: Can CGPA be predicted from learning-behavior patterns?

Previous Research Insights

University Records

The strongest predictors of academic performance

“Students' mid-term-exam grades are an important predictor to be used in predicting their final-exam grades“ (Yağcı 2022)

Demographics

Add only marginal value when stronger features are present

“Only admission marks and final marks are used; no socio-economic or demographic features are considered.” (Asif et al. 2017)

Learning Behavior

Often depending on verified data and features metrics that enable accurate grade prediction

“the amount of study only emerged as a significant predictor of cumulative GPA when the quality of study and previously attained performance were taken into consideration” (Plant et al. 2005)

Why the research important?

This research helps **validate insights** and **challenge** existing assumptions about **predicting grades** based solely on **self-reported learning habits.**

Data Overview

1,195
Students
Computer Science and Engineering students in Bangladesh

31
Features
Self-reported university records, study habits, technology access and background traits

2021
First Collection Year
Contains information about students who began studies between 2013-2023

Research Main Features

Academic History
Previous academic performance records

Learning Behavior
Study habits and learning patterns

Demographic
Background and personal characteristics

Methodology

Data Cleaning

Feature engeneering

Analyze features and selection

Fit Model and analyze

Compere to previous research insights

Conclusion and future work

Methods And Evaluation

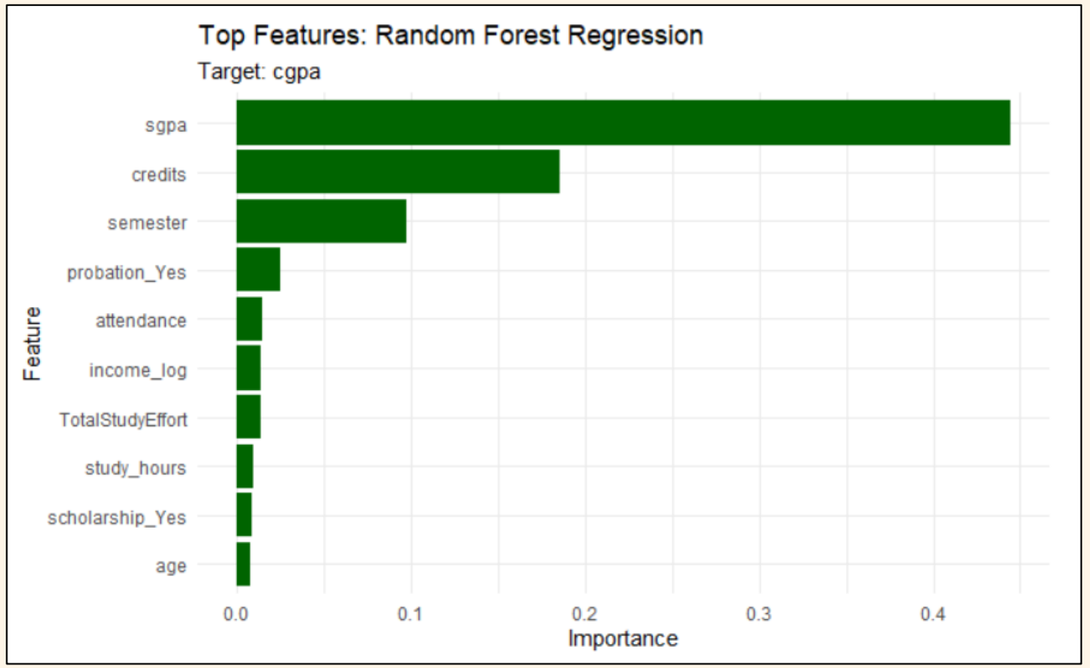
	Linear Reg.	RF Reg.	RF Classifier
Valuate Indicators	R^2, $R^2\ Adj, RMSE, MAE$	R^2 ,$R^2\ Adj, RMSE, MAE$	$F1$, $Accuracy$, $precision$, $Recall$
Target feature	CGPA (0-4)	CGPA (0-4)	CGPA_CLASS (A-D)

Results

General Question Results:

Model Type	Task Type	Metrics	Feature Importance
Random Forest	Regression	R^2 = 0.8706, RMSE = 7.4898	sgpa, credit, semester
Random Forest	Classification	F1 = 0.7365, Accuracy = 0.7368	sgpa, credits, income_log

Feature Importance:



Results (Continued)

RQ1 Results:

Model Type	Task Type	Metrics	Feature Importance
Random Forest	Regression	R^2 = 0.9245, RMSE = 0.2163	sgpa, credits, semester
Random Forest	Classification	F1 = 0.7822, Accuracy = 0.7812	sgpa, credits, semester

RQ2 Results:

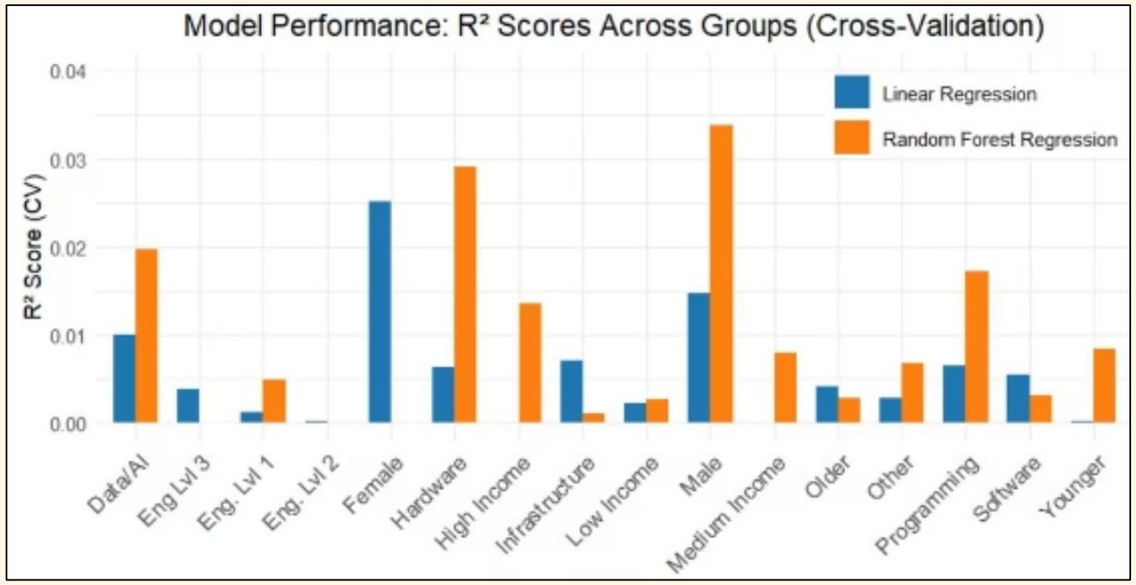
Model Type	Task Type	Metrics	Feature Importance
Linear Regression	Regression	R^2 = 0.0418, RMSE = 19.2894	english_Basic, english_Intermediate, relationship_Married
Random Forest	Classification	F1 = 0.3297, Accuracy = 0.3333	relationship_Married, english_Basic, english_Intermediate

Results (Continued)

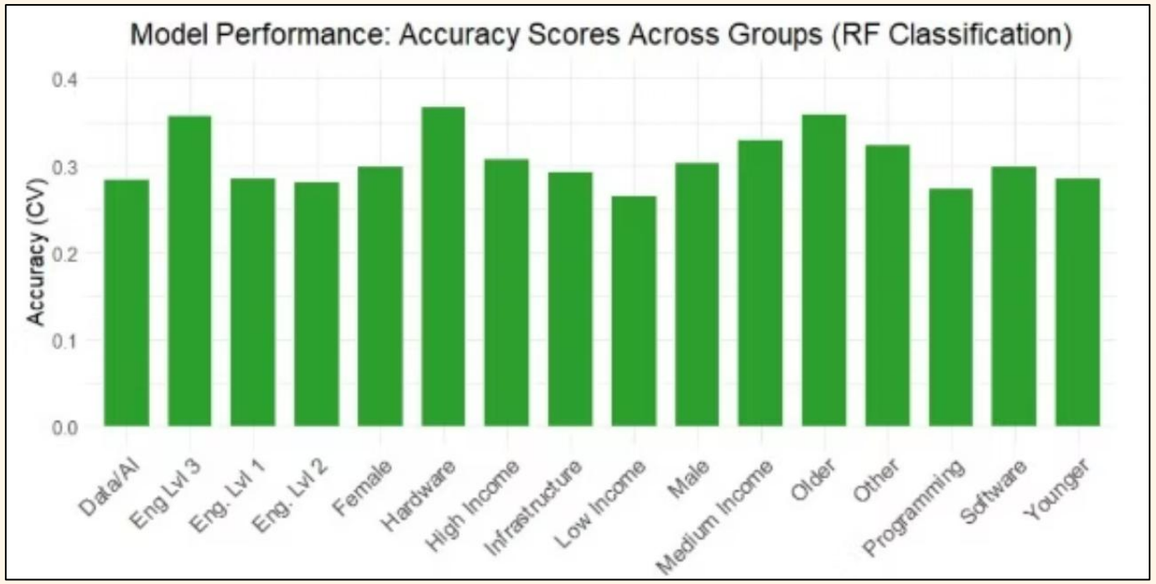
RQ3 Results:

Model Type	Task Type	Metrics	Feature Importance
Linear Regression	Regression	R^2 = 0.0008, RMSE = 19.2894	attendance, study_hours, study_freq
Random Forest	Classification	F1 = 0.3297, Accuracy = 0.3333	attendance, study_hours, study_freq

Sub-group Accuracy (CV)



Sub-group R^2 (CV)



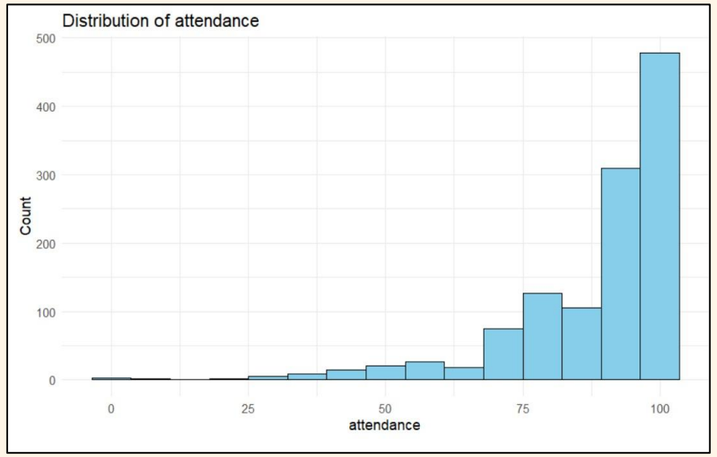
Limitations

1. Self-report bias
2. Small variance of variables (narrow range)
3. Unverified data that doesn't necessarily indicate learning quality

65%

Attendance

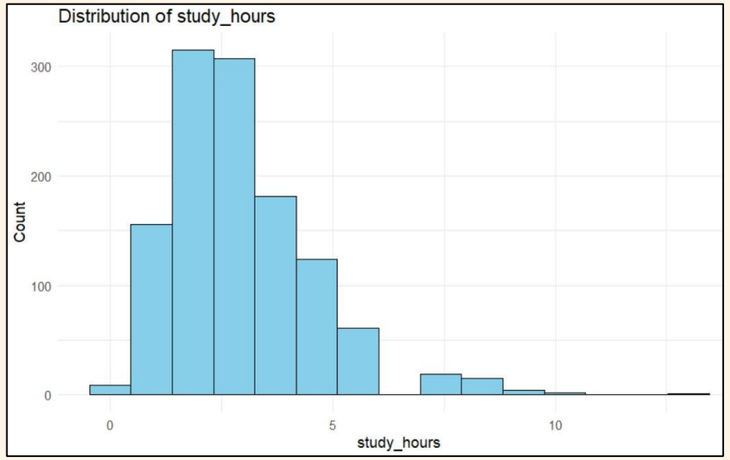
Students report attendance ≥ 90%



80%

Study Time

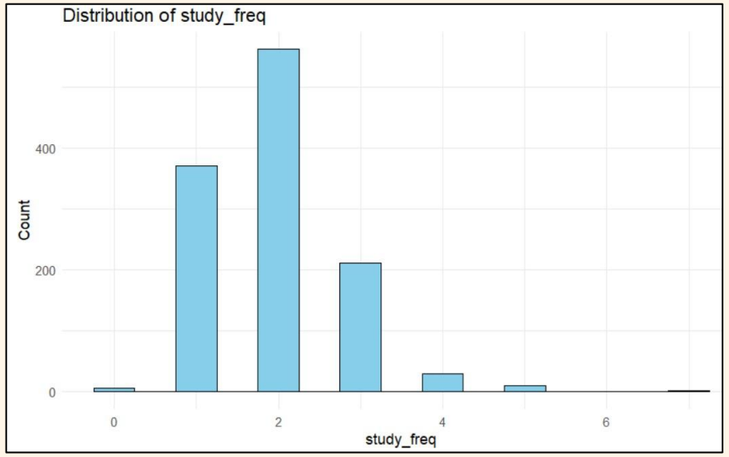
Students study between 1-4 hours per day

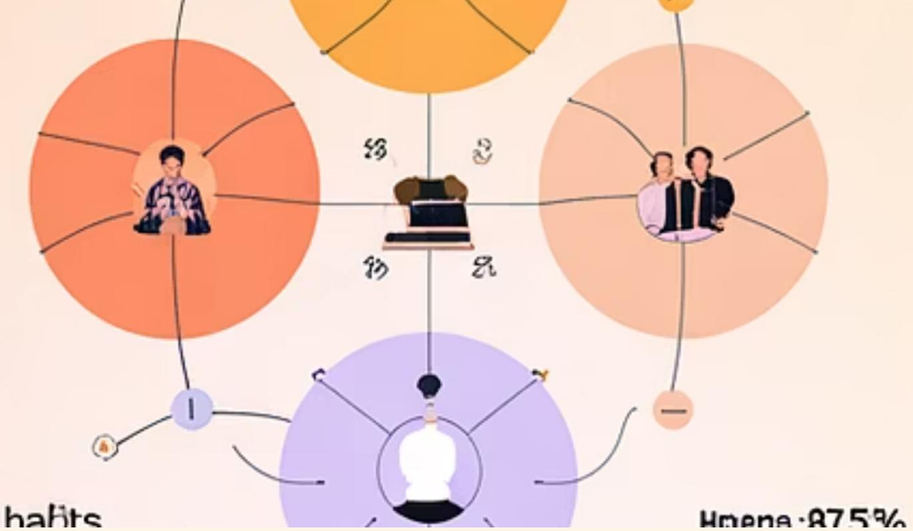


95%

Study Frequency

Report study frequency of 1-3 times





Findings

Verified University Records

Found to be the **best predictors** of final grades

Demographic Data

Found to be marginal predictors that **cannot be used alone for prediction**

Learning Habits Data

Found to be **very weak predictors** of final grades in our data

Key Insights & Conclusions

Research Questions 1 & 2

Confirmed prior findings: Academic metrics were strong predictors of CGPA; demographics alone had little to no predictive power.

Research Question 3

Accurate grade prediction based on learning habits requires **verified data** and **quality-related** metrics, **self-reports alone are not enough.**