

# Analyzing students' academic performance using educational data mining

Sazol Sarker, Mahit Kumar Paul\*, Sheikh Tasnimul Hasan Thasin, Md. Al Mehedi Hasan

Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Kazla, Rajshahi-6204, Bangladesh

## ARTICLE INFO

### Keywords:

Educational data mining  
Classification  
Academic performance  
Performance progression  
Consistent performance

## ABSTRACT

Educational Data Mining (EDM) is the process of extracting useful information and knowledge from educational data. EDM identifies patterns and trends from educational data, which can be used to improve academic curriculum, teaching and assessment methods, and students' academic performance. Thus, this study uses EDM techniques to analyze the performance of higher secondary students in Bangladesh. Three crucial categories, such as good, average, and poorly-performing students are considered for analysis. Four significant aspects of students' performance are emphasized for evaluation in this study. Firstly, predicting students' academic final examination performance in terms of internal college examination. Secondly, identifying all subjects' impact on classifier performance. Thirdly, examining students' performance progression during their studies and relating with subject-wise improvement or degradation. Fourthly, discovering consistent patterns of students' performance based on previous internal examination performance trends. The classification result reveals the correlation between internal examination and final academic performance. In addition, it resembles the predictor subjects for academic performance. The result also highlights the consistent pattern of students' consecutive internal examinations' performance. Thereafter, college administration can take necessary supportive initiatives for poorly-performing students and encourage good-performer students to continue excelling.

## 1. Introduction

The development of a country is not possible without skilled people. One of the fundamental ways to enrich skills is through education (Maxwell, 2012). That's why the education system of a country focuses more on educating every citizen. However, in Bangladesh, a considerable percentage of college students today perform poorly on their academic examinations because of drawbacks such as low quality teaching and limited learning facilities, inadequate technical and vocational training, socioeconomic factors, lack of motivation and engagement, and significant school dropout rate (Kono et al., 2018). As a result, the dream of building a skilled nation is at risk. If we can identify the reasons behind poor academic performance, then by addressing those issues, both students and the nation can benefit. Data mining can help us identify the reasons behind poor academic performance and can also provide valuable insights into current performance

trends and potential improvements. Nowadays, huge amounts of students' data are available. Extracting certain pattern or knowledge from those data is called data mining (Han et al., 2022). Educational data mining is one of the major fields of data mining, which can be categorized as prediction, relationship mining, clustering, discovery within models, and distillation of data for human judgment (Baker, 2010). This study aims to explore students' academic performance from various aspects focusing on current performance patterns and further performance improvement scopes.

In 2020, because of COVID-19 spread-out, government of the People's Republic of Bangladesh rescheduled the Higher Secondary Certificate (HSC)<sup>1</sup> examination several times and finally decided that HSC examination would not be conducted considering the mass health concern and infection issue due to the pandemic. Therefore, the government followed education experts suggestion for deciding result of more than 1.3 million HSC candidates (Star, 2021)(Tribune, 2020). Accordingly,

\* Corresponding author.

E-mail addresses: [sazolsarker1@gmail.com](mailto:sazolsarker1@gmail.com) (S. Sarker), [mahit.cse@gmail.com](mailto:mahit.cse@gmail.com) (M.K. Paul), [thasinsheikh1998@gmail.com](mailto:thasinsheikh1998@gmail.com) (S.T.H. Thasin), [mehedi\\_ru@yahoo.com](mailto:mehedi_ru@yahoo.com) (M.A.M. Hasan).

<sup>1</sup> Higher Secondary Certificate (HSC) is a public examination in Bangladesh conducted by the Boards of Intermediate and Secondary Education under the Ministry of Education. HSC is the continuation of the Secondary Education Courses and it precedes the Tertiary Education governed by the universities. Class XI-XII (intend to sit for HSC examination) roughly covers the 16-18 age group in the context of Bangladesh.

the government decided that the HSC result would be published based on the Junior School Certificate (JSC)<sup>2</sup> and Secondary School Certificate (SSC)<sup>3</sup> results (Alo, 2020). This decision is to be appreciated considering the COVID-19 pandemic situation as it was quite effective for reducing academic session jam. Academic session jam refers to a situation where the regular academic calendar or session of an educational institution gets disrupted or delayed due to various reasons such as strikes, protests, natural disasters, administrative issues, or any other factors that prevent the smooth functioning of the academic year (The Finance Today, 2023). For this reason, this study has analyzed to find out what could be the scenario if the HSC examination had been conducted by proposing GPA (Grade Point Average) calculation techniques that are comparable with board GPA.<sup>4</sup> In a word, comparison among two proposed GPA calculations with board GPA in terms of academic performance relevancy clearly resembles the impact of COVID-19 pandemic.

In this work, we worked on predicting students' HSC examination performance and proposed two different approaches for the calculation of HSC GPA using their internal college examination marks. Then we compared the generated results with the education board provided GPA using different classification methods. Also, this study aims to determine which subjects have less or more impact on students' results (obtained GPA). For our other objectives, we analyzed the data to find out in which subject a student performed highly or poorly. Additionally, we identified subjects for each student that can greatly assist in performance improvement with minor effort. Besides, an initiative has been established to distinguish all subjects as a performance booster, degraded, or at-risk based on the number of affected students. This aspect of performance observation may prove significantly helpful in reducing student failure in future examinations. Furthermore, we attempted to identify students who had performed consistently in their internal examinations much earlier so that they could improve their upcoming performance with proper guidance. Eventually, academic performance-related useful information, patterns, trends, and insights provided by the various aspects followed by the research goals of this study can surely guide students to understand current performance, lackings in performance, and alert them earlier to improve performance. The authorities can also act effectively and update academic strategies accordingly.

### 1.1. Research objectives

This study aims to analyze the performance of students studying a two-year HSC program of humanities division at a college in Bangladesh. The objective is to observe the continuous academic performance of both student groups and individual student precisely. Also, the focus is on to provide useful knowledge and patterns regarding students' academic performance to the concerned teachers and college authorities. This study seeks to accomplish four research objectives as follows.

- *Firstly*, several classification models are generated to predict the performance of students at the end of the HSC degree. To build these classifiers, only academic internal examination marks of first and

second year within college are used. No categorical attributes (socioeconomic and demographic factors) of students are considered. Classification models provide a summary of students' academic performance. This approach will enable the college authority to keep track of institutional overall performance in HSC examination based on internal examination subject marks only.

- *Secondly*, we aim to derive predictor subjects and their impact that can serve as effective indicators for students' performance prediction in HSC examination. As a result, we will be able to identify good, average and poor performer students and support at-risk students as soon as possible. This can be accomplished by using human interpretable classifiers that provides simple visual classification outcomes.

According to (Asif et al., 2017), using a random classifier other than decision tree will not be a good choice to find predictor subjects, because the prediction capacity and interpretability of a classifier model may need to be traded off. Therefore, we used Decision Tree (DT), a classifier algorithm explained in section 3 for this study. The decision tree is the most simple model that gives highly interpretable decision tree graph outcomes for human visualization based on different statistical feature selection criteria (Asif et al., 2017).

DT with different impurity measures, such as information gain, gini index and accuracy, may sometimes lead to biased results (Raileanu & Stoffel, 2004). According to (Han et al., 2022), information gain gives bias result in favor of multivalued attributes in dataset. Gini index has tendency to favor tests that result in equal-sized and pure partitions (Liu et al., 2018). Besides, it can not handle large number of classes and gets biased to multivalued attributes as well. Due to imbalanced class distribution in dataset, DT with accuracy produces biased trees and favors the majority class (Han et al., 2022).

Considering the circumstances, we suggested an approach to identify predictor subjects and their impact. We did the task by observing the DT graph node outcomes using the concept of voting. This greatly paves the way to compare classifier performance and student performance-subjects correlation between the proposed and board GPA approach. This helps to reach an important conclusion about efficient attribute selection and attribute-label association. Furthermore, this approach opens up an opportunity for non-technical individuals to comprehend our second research objective. It allows them to easily interpret and relate student performance with subjects without requiring prior knowledge of machine learning.

- *Thirdly*, through our proposed approach we investigated how students' academic performance progresses over the two-year academic study period. Using average marks in internal examination as performance baseline, the subject-wise continuous progression of each student can be determined. The subjects that boost overall GPA are defined as booster subjects and the subjects that are responsible for decreasing the GPA are called degrader subjects. Again, failure-prone subjects are addressed as at-risk subjects. This way every student's progression through internal examination can be properly examined. Besides, determining booster subjects and degrader subjects based on the number of affected students, few of them can be shortlisted to improve subject-wise academic performance. The proposed continuous progression model presents both highly interpretable and visually clear outcomes that even non-technical persons like college teachers and students can understand easily. This paves the way to closely monitor students and detect at-risk subjects. Consequently, students' result degradation can be significantly reduced with immediate supportive measures.
- *Fourthly*, consistent performance pattern detection during academic periods based on two consecutive previous internal examination performance trends can be helpful. It can remarkably help in detecting the bright and poor performer students (e.g., struggling, at-risk) well in advance of their upcoming internal examinations. It can

<sup>2</sup> Junior School Certificate (JSC) is a public examination taken by students in Bangladesh after the successful completion of eight years of schooling. It is introduced in 2010. It is followed by the Secondary School Certificate (SSC) examination.

<sup>3</sup> Secondary School Certificate (SSC) is a public examination in Bangladesh conducted by the Boards of Intermediate and Secondary Education under the Ministry of Education. SSC is held based on the books of classes IX and X, which are usually the same.

<sup>4</sup> In this paper, board GPA refers to the Grade Point Average (GPA) calculation of HSC examination, during COVID-19 pandemic, taken by the Boards of Intermediate and Secondary Education. The corresponding grading system calculation, used during COVID-19 pandemic, is available at <https://en.prothomalo.com/youth/education/hsc-results-of-2020-in-january-after-issuance-of-ordinance-dipu-moni>.

preciously help in triggering timely intervention for performance improvement through special teaching and assessment methods.

However, the objectives can be summarized as follows:

- **Objective 1:** Predicting students' final academic performance based on internal exam performance using several classifiers. Additionally, conducting a comparative analysis between proposed GPA and board GPA in terms of relevancy with internal examination performance.
- **Objective 2:** Identifying predictor subjects and their impact on students' performance prediction, besides comparing between proposed GPA and board GPA in terms of classification accuracy and feature selection efficiency.
- **Objective 3:** Investigating students' performance progression during academic two-year period.
  - i) Finding individual student's GPA booster subjects and degrader subjects.
  - ii) Discovering individual student's at-risk subjects.
  - iii) Categorizing all subjects into at-risk subjects, performance booster, and degrader subjects.
  - iv) Pinpointing present performance statistics and performance booster subjects with low effort for individual student performance improvement.
- **Objective 4:** Finding a consistent pattern of students' performance based on previous consecutive internal examinations' performance trends.

The remainder of the paper is organized into several sections. In section 2, the work done by distinguished researchers considering educational data are discussed. The classifiers which we used for classification purposes are illustrated in section 3. In section 4, the dataset used and the workflow of the proposed approach are discussed. The detailed performance analysis of our work is investigated in section 5. The limitations and possible future work scopes of this study are discussed in section 6. In section 7, the theoretical and pedagogical implications of this study are provided briefly. Finally, section 8 concludes our work.

## 2. Literature review

A nation can not prosper without educating its citizens (Edu, 2023). Because of this, a lot of scholars have worked hard to determine how they might predict students' performance using a variety of factors. (Asif et al., 2017) addressed three major questions in their work. For this, they used university exam marks from the first two years. At an early stage, they predicted students' performance, relating index courses that can serve as pointers of low or high performance, and relating typical signs of progress. They used ten classifiers to predict performance. In their study, they described that there is a reasonable relationship between students' final examination results with their previous internal college examination marks. Also, they suggested some impactful subjects. They claimed that by giving more emphasis on these courses, students could perform better on examinations.

(Mishra et al., 2014) used data mining for students' performance prediction. Finding a comparatively better prediction model and the impact of various attributes on students' performance were the main objectives. The dataset contained marks, socio-economic, and demographic data of the students. Different classification techniques were used to build a performance prediction model based on students' semester marks, social integration, academic integration, and emotional skills. J48 and Random Tree were applied to predict the third semester performance of students, with Random Tree found to be more accurate than J48. The authors also identified attributes that influenced students' third semester performance.

The performance of university students was predicted at an earlier time by (Meghji et al., 2023) using 291 university students data. They

also made an effort to find out how the courses affected the students' performance. They used decision trees for their purpose. Additionally, they proposed a framework for segmenting students based on their performance, which can be used to create practical educational policies.

(Baradwaj & Pal, 2011) delved into students' academic performance at the end of the semester and identified dropout students. They used some attributes that are related to students' semester results like previous semester marks, performance, assignment, attendance, last semester marks, and so on to predict performance. They used decision tree for prediction.

In a recent work (Tasnim et al., 2019) proposed a threshold based approach that can be used to identify drop out students' using some attributes like student's sex, travel time, current health status, number of school absences, and so on. They showed that their threshold based approach gives better classification results compared to naive bayes, logistic regression, and support vector machine.

(Golding & Donaldson, 2006) examined the correlation between students' performance (e.g. GPA) and predictor courses in the Bachelor of Science and Information Technology (BSCIT) program. They examined that lectures that ensure foundation concepts are tutored and understood by students can be one of the major factors for success. Though they also used demographic factors for predicting performance, they found no relation between them.

In another work (Huang & Fang, 2013) predicted students' final examination marks using various types of mathematical models. They used a dataset collected from 323 undergraduate students in various semesters. In their dataset, they included marks in various subjects and also CGPA in four semesters. They concluded that for measuring average performance logistic regression performs better, and for determining individual performance support vector machine has the highest accuracy.

The study of (Mashiloane & Mchunu, 2013) focuses on using Educational Data Mining to predict first year students' success or failure based on first semester marks at the university. Predictive models were obtained from the training data set using J48 classifier, Decision Table, and Naive Bayes. J48 was found to be the better performing classifier. The findings also suggest that first semester marks can be an effective factor for the prediction of first year Computer Science final results, potentially helping to identify at-risk students for early intervention.

(Kaunang & Rotikan, 2018) conducted a comparative analysis between Decision Tree and Random Forest predicting the academic performance of students using WEKA. The dataset contains students' demographics, previous GPA, and family background information which were used to construct students' academic performance prediction models. This study's main objectives were to find a better prediction model and factors that affect students' learning behavior.

(Czibula et al., 2019) analyzed students' academic performance prediction using supervised and unsupervised learning methods. The authors proposed a new classification model, namely *S PRAR* (Students Performance Prediction using Relational Association Rules), which used relational association rules to predict students' final results in academic disciplines. Experiments on real academic datasets showed that *S PRAR* outperformed other existing students' performance predictors.

(Bujang et al., 2021) indicated the urgent need for predictive analytics in higher education and the challenges of using machine learning to predict student grades. The authors presented a comprehensive analysis of six machine learning techniques comparing their accuracy in predicting student grades using real student course data. Additionally, they proposed a multi-class prediction model to address imbalanced datasets. The model integrates with Random Forest to achieve significant f-measure improvement. This model showed promising results for enhancing predictive performance in imbalanced multi-classification for student grade prediction.

(Yang & Li, 2018) predicted students' performance and suggested a few elements that are crucial for advancement. They demonstrated how students could improve more by concentrating on those characteristics.

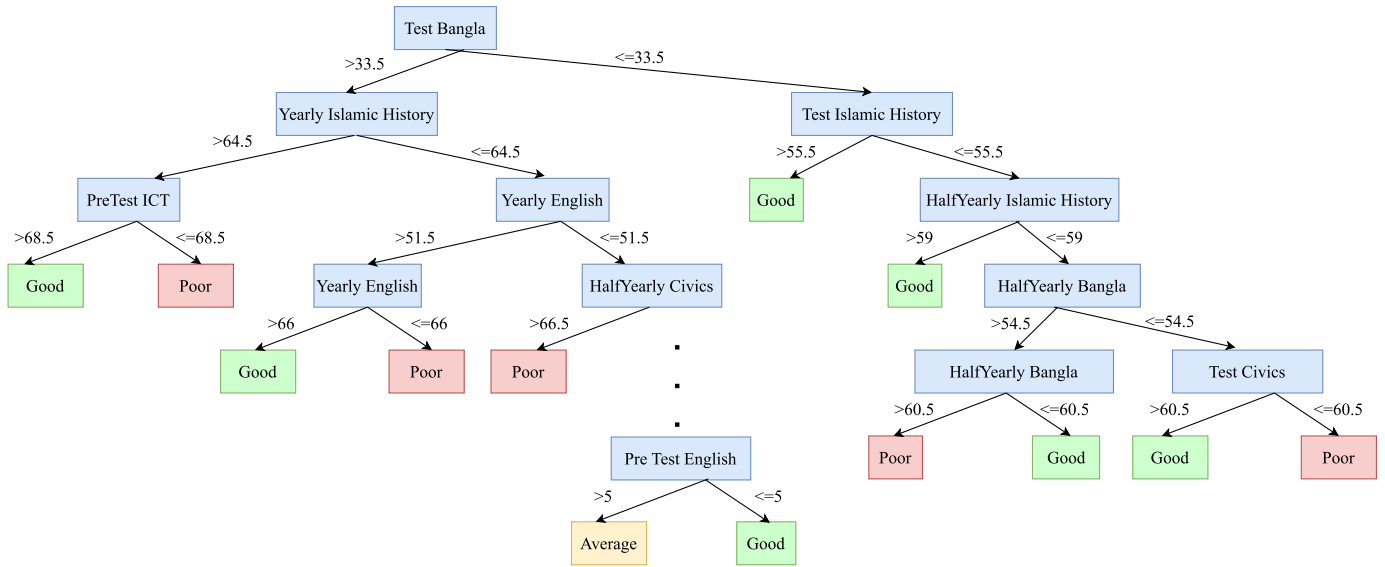


Fig. 1. Decision tree with Accuracy (Board GPA).

According to Guzmán et al. (2007), a self-evaluation test could be a crucial tool for students' progression.

The previous studies show that academic marks, socioeconomic status, demographics, and other variables, might affect a student's performance. Researchers have predicted student performance using both single and multiple factors. In this paper, we have used a single attribute i.e., academic marks to predict students' performance. Since no algorithm consistently produces the greatest outcomes, as indicated by the previous studies, we have used five machine learning algorithm models that are appropriate for obtaining the objectives of this study. In addition, the majority of the previous study revolved around binary classification; but, this study introduces multi-class classification, which will enable us to more precisely determine each student's performance level. Besides, predicting the students' performance was the main concern of most of the previous studies, but this study has further addressed various performance factors for students' level of performance. Thus, this study has additionally explored subjects' impact on results and introduced individual, group-wise performance observation, consistent performance trend monitoring, which are the unique contributions of this study.

### 3. Data mining techniques

In this study, we used five machine learning algorithms for analyzing students' academic performance from various perspectives. For the first and fourth research objectives, all of the five machine learning algorithms are used. Then we selected one of these models based on the highest classification performance to extract performance related further insights. In the case of the second research objective, we only used Decision Tree with various statistical criteria to visualize the DT model prediction outcomes as a bundle of subjects and conditions. In this way, subjects and their impact on academic results are extracted from analytical results with the help of machine learning algorithms. Besides, this study has proposed a model to classify individual and group wise student performance in the third research objective without using any help from machine learning algorithms.

#### 3.1. Decision tree

Decision tree is a popular machine learning algorithm that is used for classification problems. It is a tree-like structure where each node represents a decision, and the branches represent the possible outcomes of that decision. The goal is to use the features of the dataset to create a tree that can accurately predict the target variable (Tangirala, 2020).

Decision tree works by recursively splitting the data into subsets based on the values of the features. At each step, the algorithm selects the feature that best separates the data into the purest subsets, meaning that the subsets contain mostly one class or category. This process continues until all the data is classified or a stopping criterion is met (Asif et al., 2017). There are several methods for selecting the best feature to split the data, such as Information Gain, Gini Index, and Accuracy.

Information Gain measures the reduction in entropy or uncertainty in the target variable when a feature is used to split the data. Features that provide the most information gain are selected first (Han et al., 2022). Gini Index measures the impurity of the target variable in a subset. Features that reduce the Gini Index mostly are selected first (Han et al., 2022). Accuracy simply selects the feature that leads to the highest accuracy on the training data (Han et al., 2022).

Some decision tree graphs generated by using Accuracy, Gini Index, and Information Gain criteria concerning board GPA analysis are shown in Figs. 1, 2, and 3 respectively. These figures represent three overfitted decision trees that resemble complex trees with many levels. The trees result in less interpretability, poor generalization and classification, and a negative impact on model performance. On the other hand, Fig. 4 for proposed GPA-1 analysis and Fig. 5 for proposed GPA-2 analysis represent a much better interpretation of the features and better classification performance of the machine learning model.

#### 3.2. K-nearest neighbors

K-Nearest Neighbors (K-NN) is a simple and intuitive classification algorithm. It works by finding the 'K' training data points (neighbors) in the dataset that are closest in terms of distance to the new data point being classified (Kramer & Kramer, 2013). The Euclidean distance is often used to calculate the neighboring distance. It is the smallest distance between any two neighbors and is always a straight line (Leung et al., 2013). The majority class label among these K-Nearest Neighbors is then assigned as the predicted class label for the new data point. The number of nearest neighbors,  $K = 1$ , is often used for binary classification tasks (Kramer & Kramer, 2013). But, this study focuses on multi-class classification and thus increasing value of K is required for better performance which is similar to the work by (Yuan et al., 2008). In this study, the number of nearest neighbors is fixed at  $K = 5$ . However, K-NN can be sensitive to noisy data, and the choice of 'K' impacts the model's performance. Also, the prediction can be computationally expensive, especially for large datasets, as it requires calculating distances for each new data point against all training examples (Sun & Huang, 2010).

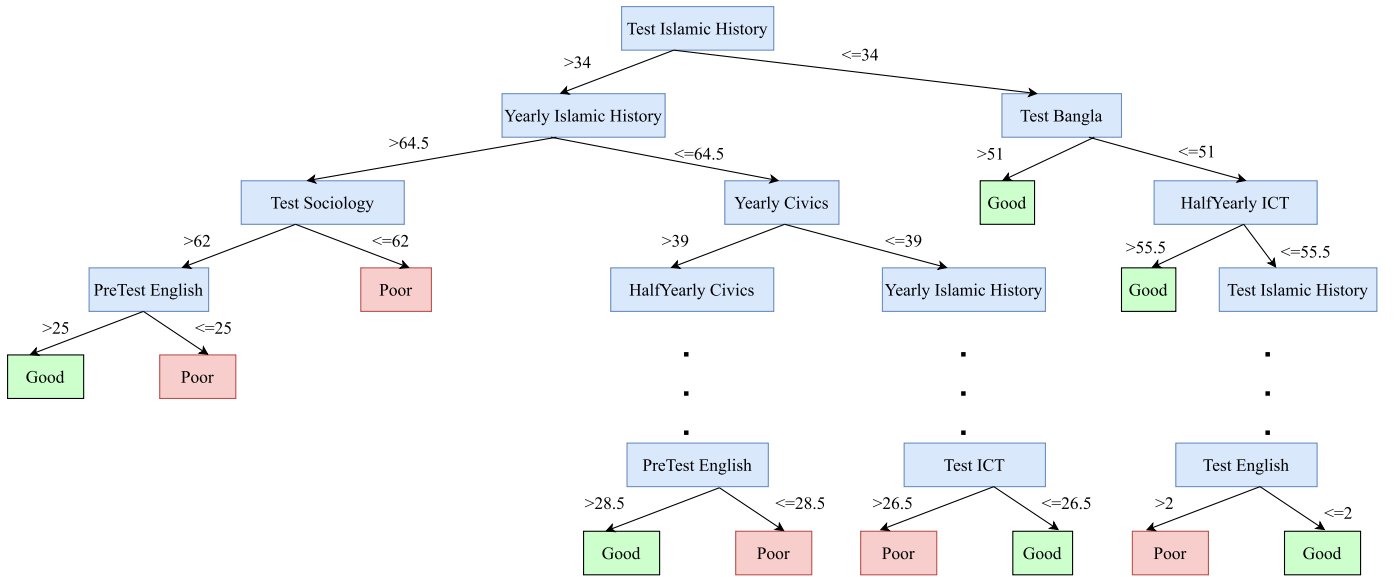


Fig. 2. Decision tree with Gini Index (Board GPA).

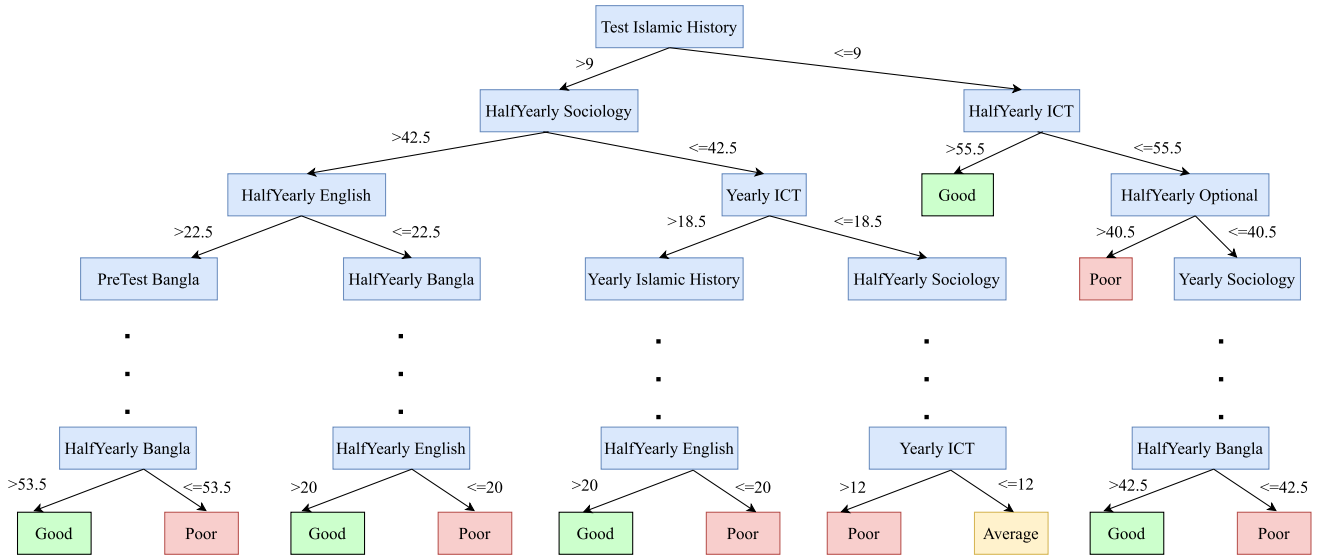


Fig. 3. Decision tree with Information Gain (Board GPA).

### 3.3. Naïve Bayes

Naive Bayes is a probabilistic machine learning algorithm used for classification tasks (Duda et al., 1973). The algorithm is based on Bayes theorem and assumes that all features in the data are independent of each other (Friedman et al., 1997). The algorithm calculates the prior probabilities of each class based on the frequency of each class in the training set. It also calculates the conditional probabilities of each feature given each class based on the frequency of each feature in the training set. When a new instance is presented, the algorithm calculates the probability of the instance belonging to each class based on the prior and conditional probabilities. The class with the highest probability is then predicted as the class of the new instance (Ren et al., 2009).

Bayes theorem finds the likelihood of an occurrence given the probability of another event that has already occurred (Efron, 2013). Bayes theorem can be defined by equation (1).

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

where,

$P(C|X)$  : posterior probability of target class

$P(X|C)$  : likelihood that is probability of predictor of given class

$P(C)$  : prior probability of class

$P(X)$  : prior probability of predictor of class

### 3.4. Neural networks

Neural Networks are a type of machine learning algorithm inspired by the structure and function of the human brain. They consist of a network of interconnected nodes, called neurons, which are organized into layers (Beale & Jackson, 1990). Each neuron receives input from other neurons of other layers, processes the information, and produces an output that is transmitted to other neurons in the network. The basic structure of a neural network consists of an input layer, one or more hidden layers, and an output layer (Beale & Jackson, 1990). The input layer receives the input data, and the output layer produces the output. The input data is processed by the hidden layers, which also extract features for use in forecasting or making judgments (Beale & Jackson, 1990).

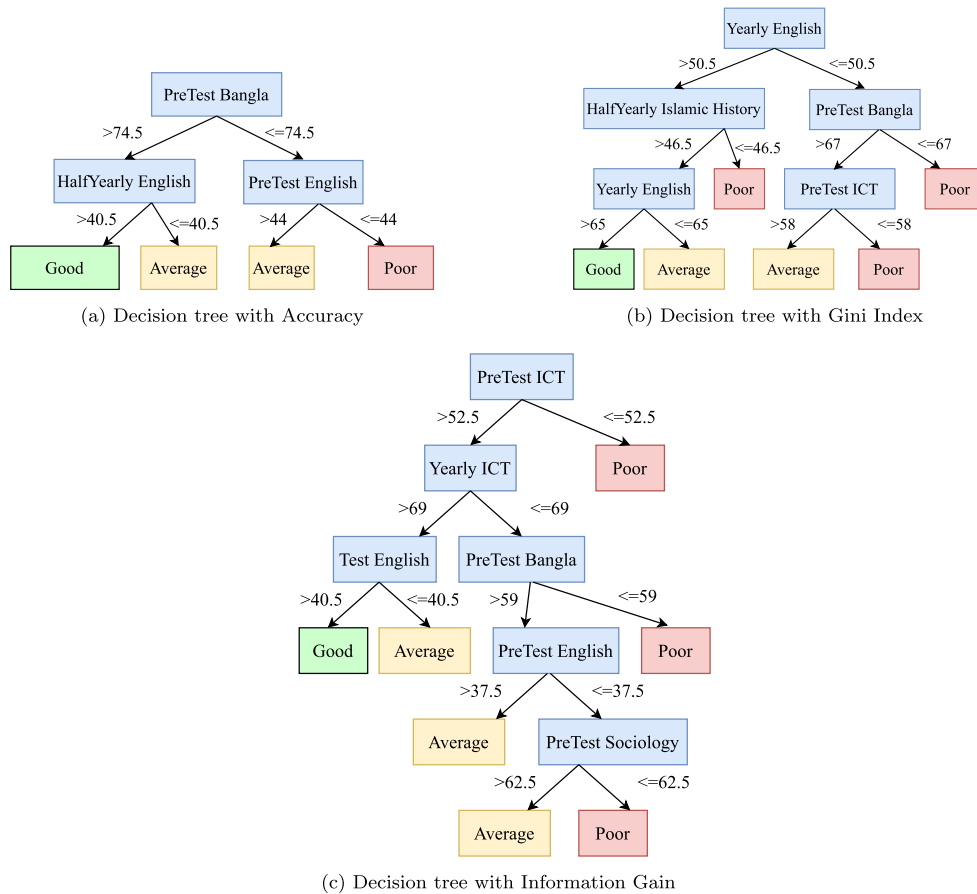


Fig. 4. Decision tree with (a) Accuracy, (b) Gini Index, and (c) Information Gain (Proposed GPA-1).

Weighted links that control the strength of the connections between the neurons in a neural network connect the neurons. During training, the weights of these connections are adjusted to help the network learn to recognize patterns and generate accurate predictions (Müller et al., 1995).

### 3.5. Random forest

Several decision trees are used in the Random Forest ensemble learning technique to get more precise predictions. Both classification and regression tasks can be accomplished using this well-known and effective machine learning technique (Tangirala, 2020). The fundamental principle of Random Forest is to build many decision trees, each trained on a random subset of the training data and a random selection of characteristics (Tangirala, 2020). This randomization aids in lowering overfitting and enhancing the model's generalization capabilities. During training, each decision tree is grown using a modified version of the CART (Classification and Regression Trees) algorithm. At each node of the tree, a subset of the features is randomly selected, and the best split is chosen based on the Gini impurity or entropy of the data. To make a prediction for a new instance, the algorithm passes the instance through each decision tree in the forest and aggregates the results. For classification tasks, the class with the highest frequency among the predictions is chosen as the final prediction.

## 4. Dataset and experimental methodology

### 4.1. Dataset

Dataset used for this study comprises students' marks in a two-year higher secondary degree of humanities division of an intermediate col-

Table 1

Dataset characteristics.

Dataset name	Attribute type	No. of attributes	No. of instances
College dataset	Integer	28	309
Synthetic dataset	Integer	28	1000

lege in Bangladesh. Besides, a randomly generated synthetic dataset with better class wise data distribution is also used for first research objective. This study contains marks data of four internal academic examinations (Half Yearly, Yearly, PreTest, Test), each with seven subjects marks data of 309 college students of academic batch (2018-2019) and data of 1000 students in synthetic dataset shown in Table 1. The dataset consists of the attributes listed in Table 2 that relate to internal examination subject marks and internal examination GPA, board GPA, and proposed GPA.

The interval of marks is divided into three categories: Good (60%-100%), Average (50%-59%), and Poor (0%-49%). The class-wise student distribution statistics are shown in Table 3 for the college dataset and the synthetic dataset in Table 4. Fails are counted as poor performance and students dropping out are not investigated in this work. Every student's marks in examinations are calculated in between 0 (Worst) to 100 (Best) for the college dataset and in between 30 (Worst) to 90 (Best) in the synthetic dataset. Selected mark range in the synthetic dataset has been proved to successfully generate random marks that resulted in the desired multi-class dataset while several other possible ranges failed to do so. All data mining methods have been implemented using Rapid-Miner software and manually coded at Google Colaboratory to obtain the required evaluation metrics.

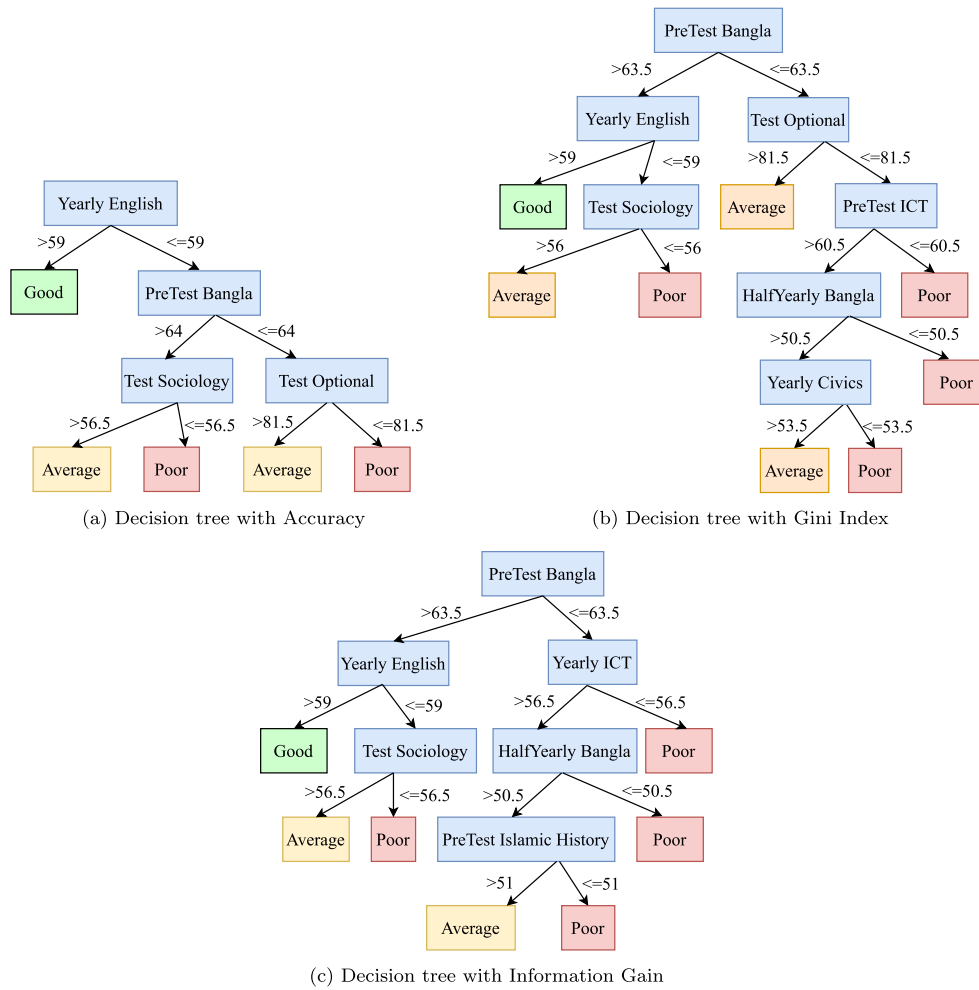


Fig. 5. Decision tree with (a) Accuracy, (b) Gini Index, and (c) Information Gain (Proposed GPA-2).

Table 2

Attributes in dataset.

Attribute	Description
HalfYearly English	Half yearly examination subject based on English literature
Yearly English	Yearly examination subject based on English literature
PreTest English	Pretest examination subject based on English literature
Test English	Test examination subject based on English literature
HalfYearly Bangla	Half yearly examination subject based on Bangla literature
Yearly Bangla	Yearly examination subject based on Bangla literature
PreTest Bangla	Pretest examination subject based on Bangla literature
Test Bangla	Test examination subject based on Bangla literature
HalfYearly ICT	Half yearly examination subject based on ICT basics
Yearly ICT	Yearly examination subject based on ICT basics
PreTest ICT	Pretest examination subject based on ICT basics
Test ICT	Test examination subject based on ICT basics
HalfYearly Civics	Half yearly examination subject based on study of citizenship and governance
Yearly Civics	Yearly examination subject based on study of citizenship and governance
PreTest Civics	Pretest examination subject based on study of citizenship and governance
Test Civics	Test examination subject based on study of citizenship and governance
HalfYearly Sociology	Half yearly examination subject based on Sociology
Yearly Sociology	Yearly examination subject based on Sociology
PreTest Sociology	Pretest examination subject based on Sociology
Test Sociology	Test examination subject based on Sociology
HalfYearly Islamic History	Half yearly examination subject based on Islamic History
Yearly Islamic History	Yearly examination subject based on Islamic History
PreTest Islamic History	Pretest examination subject based on Islamic History
Test Islamic History	Test examination subject based on Islamic History
HalfYearly Optional	Half yearly examination subject (chosen 1 out of few)
Yearly Optional	Yearly examination subject (chosen 1 out of few)
PreTest Optional	Pretest examination subject (chosen 1 out of few)
Test Optional	Test examination subject (chosen 1 out of few)
Board GPA	HSC board result (Alo, 2020)

**Table 3**  
Statistics of class-wise student distribution in college dataset.

Analysis label	College dataset		
	No. of students (Good)	No. of students (Average)	No. of students (Poor)
Board GPA	152	40	117
Proposed GPA-1	4	13	292
Proposed GPA-2	5	13	291

**Table 4**  
Statistics of class-wise student distribution in synthetic dataset.

Analysis label	Synthetic dataset		
	No. of students (Good)	No. of students (Average)	No. of students (Poor)
Board GPA	366	422	212
Proposed GPA-1	610	372	18
Proposed GPA-2	538	421	41

#### 4.2. Experimental methodology

In the time of COVID-19 pandemic in 2019, internal college examinations (Half yearly, Yearly, PreTest, Test) were already conducted in Bangladesh, but HSC-2020 could not be conducted (World Health Organization, 2019). The government had to reschedule the examinations multiple times but led to the postponement of the HSC examinations to ensure the safety and well-being of students considering the outbreak of COVID-19 (Star, 2021). Due to the uncertainty because of the pandemic and the challenges of conducting traditional examinations, the education boards in Bangladesh opted for alternative evaluation methods. Experts sought a universal standard to evaluate all students on the same scale. They suggested evaluating HSC results based on JSC and SSC results to prevent academic session jams among multiple batches during emergencies like the COVID-19 pandemic (Tribune, 2020). Therefore, HSC result was calculated as 25 percent of the JSC result and 75 percent of the SSC result, and the education board had issued an ordinance as legal support for this result (Alo, 2020).

We observed that our concerned college had a 100% pass rate considering the board GPA. However, a considerable portion of students' internal examination performance was poor and the students might fail in HSC examination. This points us towards the significant impact of COVID-19 on HSC results. We, therefore, seek an answer to a question: What if the HSC examination had been conducted in normal circumstances, what should be the results? For this reason, we investigated the students' performance in internal academic examinations as well as in HSC GPA, addressing the first and second research objectives of this study. Thus, three analysis cases arise concerning board GPA, proposed GPA-1, and proposed GPA-2.

Five classification algorithms with different criteria were used to predict intermediate college students' HSC examination performance at the end of two-year HSC degree. This finds the answer to our first research objective that seeks classifier prediction performance of obtained HSC GPA in terms of internal examination performances.

Table 3 presents class wise distribution of students' performance in college dataset and Table 4 displays the same for synthetic dataset. Class 'Good' has the most students in college dataset and class 'Average' has the most students in synthetic dataset for board GPA analysis. Again, for proposed GPA-1 and proposed GPA-2 analysis, class 'Poor' has the most students in college dataset and class 'Good' has the most students in synthetic dataset. 5-fold cross-validation technique is used to evaluate the performance of the classification models. Preliminary analysis steps of this study include data pre-processing, proposed GPA calculation, multi-class label generation, and building classifier models shown in Fig. 6.

First research objective aims to find students' internal examination performance patterns and relate to academic final performance using several classifiers. Three comparable analysis scenarios arise because this study proposes two candidate GPA based on central tendency (mean and weighted sum) criteria along with board GPA. When calculating HSC subject-wise marks, the proposed GPA-1 technique uses the average of four internal examination subjects' marks. On the other hand, proposed GPA-2 approach implies the weighted sum of subject marks. Then subject marks are converted to equivalent GPA using standard GPA mapped to a predefined mark range. The subject-wise marks in HSC examination are calculated based on internal examination subject-wise marks using two different formulas. Equations (3) and (4) are used for proposed GPA-1. Equations (5) and (6) are used for proposed GPA-2 generation. In this way, HSC GPA is proposed based on internal examination performance. First research objective uses evaluation metrics such as accuracy, weighted F1-score, and Cohen's kappa to predict the class of the student performance with reasonable precision.

Second research objective of this study targets to find out academic performance and subjects relationship in each of the three mutually comparable analysis cases. This paves a way for non-technical person to understand the research outcomes because it presents more generalized and human interpretable outcomes. It is possible with the help of decision tree graphs shown in Figs. 1, 2, 3, 4, and 5. These obtained decision trees selected a few of the internal examination subjects to classify students according to performance class. Thereby represent correlation between subjects and students performance. Using only one impurity criterion such as information gain, gini index or accuracy can not be a wise choice as each of them behaves biased towards the dataset's majority class Han et al. (2022). That's why, to reduce biased outcomes, a novel approach is used to decide predictor subjects using an outcome voting technique within all impurity criteria which interprets DT graph nodes from a special angle.

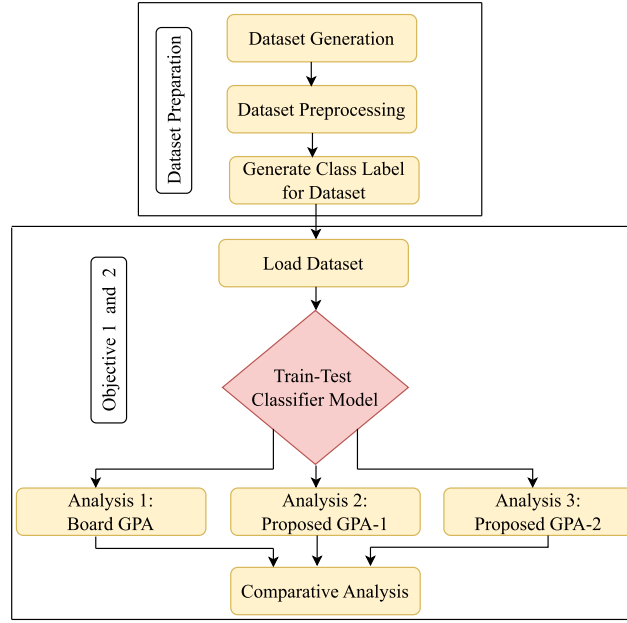
Third research objective focuses on finding every student's performance in terms of subjects. Thus categorizes all subjects as performance boosters, degraders, and at-risk subjects. It also presents a summarized view of each subject and the number of affected students by it as booster, degrader, as well as at-risk subjects. Finally, this objective points out every student's current academic performance and suggests low-effort performance enhancer subjects to improve results. Besides, booster subjects are defined as subjects that assist in performance increase, degrader subjects as performance reducers, and at-risk subjects as already failed or fail-prone subjects. At-risk criteria are fixed for subject marks less or equal to 40, whereas the passing mark baseline is 33.

Fourth research objective focuses on finding consistent student performance using previous internal examination performance trends. Consequently, it results in early detection of upcoming internal examination's most probable estimated performance. Five classification models are used for this purpose and performance has been evaluated using 5-fold cross-validation. This study proposes a novel approach that aims to find a correlation between half yearly examination subject marks and yearly examination GPA. The classification result resembles a consistent performance pattern of students in consecutive two internal examinations - half yearly and yearly, yearly and pretest, pretest and test. Since the ultimate goal is to perform excellently in upcoming internal and HSC examinations, this objective can provide very useful insight, such as forecasting performance and detecting continuously static performer students who are either consecutive good, consecutive average, or consecutive poor performers. First and second research objectives of this study include a comparison of three GPAs shown in Fig. 6 and their corresponding formulas are covered in the equations (2) - (6).

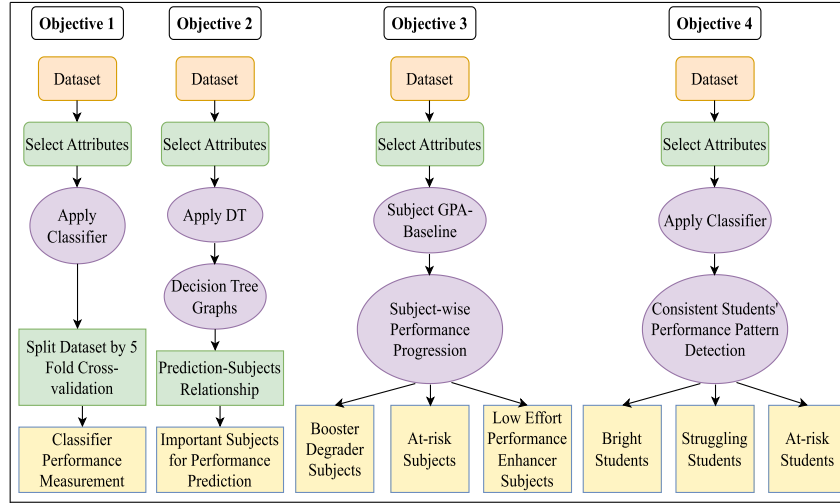
##### 1. Board GPA (Alo, 2020)

$$HSC\ GPA = (JSC\ GPA \times 0.25) + (SSC\ GPA \times 0.75) \quad (2)$$

Here, *JSC GPA* is obtained GPA in JSC examination  
*SSC GPA* is obtained GPA in SSC examination



(a) Dataset generation and process block diagram of research objective 1 and 2



(b) Research objectives (in details)

Fig. 6. Workflow of proposed methodology.

## 2. Proposed GPA-1

$$HSC\ Sub_i = \frac{(HFSub_i + FSub_i + PreSub_i + TestSub_i)}{n_1} \quad (3)$$

$$HSC\ GPA = \frac{\sum_{i=1}^n HSC\ Sub_i}{\text{Total No. of Subjects, } n} \quad (4)$$

## 3. Proposed GPA-2

$$HSC\ Sub_i = (HFSub_i \times 0.1) + (FSub_i \times 0.20) + (PreSub_i \times 0.30) + (TestSub_i \times 0.40)$$

$$HSC\ GPA = \frac{\sum_{i=1}^n HSC\ Sub_i}{\text{Total No. of Subjects, } n}$$

Here,  $HSC\ Sub_i$  is HSC  $i^{th}$  subject mark

$HFSub_i$  is Half Yearly  $i^{th}$  subject mark

$FSub_i$  is Yearly  $i^{th}$  subject mark

$PreSub_i$  is Pretest  $i^{th}$  subject mark

$TestSub_i$  is Test  $i^{th}$  subject mark

Number of internal examinations,  $n_1 = 4$

Number of total subjects,  $n = 7$

## 5. Results analysis

This section presents the experimental results of this study starting with the first research objective and ending with the fourth. Formulas of evaluation metrics for classifier performance measurement used in this study are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (11)$$

**Table 5**  
Classifier prediction accuracy.

Classifier	College dataset			Synthetic dataset		
	Board GPA	Proposed GPA-1	Proposed GPA-2	Board GPA	Proposed GPA-1	Proposed GPA-2
Decision Tree with Gini Index (DT-GI)	43.98%	95.16%	94.17%	55.70%	60.20%	61.40%
Decision Tree with Information Gain (DT-IG)	50.49%	95.46%	94.51%	57.30%	62.20%	59.70%
5-Nearest Neighbour (5-NN)	56.30%	94.83%	94.82%	57.40%	73.10%	69.70%
Naive Bayes	45.65%	89.33%	89.97%	75.70%	85.80%	83.20%
Neural Network	49.16%	96.12%	95.48%	79.60%	86.50%	82.90%
Random Forest with Gini Index (RF-GI)	57.28%	96.45%	95.15%	68.30%	76.70%	77.90%
Random Forest with Information Gain (RF-IG)	55.35%	96.44%	95.15%	71.30%	73.90%	76.90%

**Table 6**  
Classifier kappa.

Classifier	College dataset			Synthetic dataset		
	Board GPA	Proposed GPA-1	Proposed GPA-2	Board GPA	Proposed GPA-1	Proposed GPA-2
Decision Tree with Gini Index (DT-GI)	0.046	0.478	0.478	0.309	0.162	0.260
Decision Tree with Information Gain (DT-IG)	0.157	0.505	0.405	0.332	0.236	0.248
5-Nearest Neighbour (5-NN)	0.260	0.487	0.593	0.319	0.425	0.407
Naive Bayes	0.159	0.423	0.423	0.608	0.697	0.675
Neural Network	0.122	0.604	0.164	0.681	0.719	0.670
Random Forest with Gini Index (RF-GI)	0.210	0.567	0.567	0.481	0.469	0.564
Random Forest with Information Gain (RF-IG)	0.175	0.529	0.529	0.534	0.401	0.543

**Table 7**  
Classifier prediction weighted F1-score.

Classifier	College dataset			Synthetic dataset		
	Board GPA	Proposed GPA-1	Proposed GPA-2	Board GPA	Proposed GPA-1	Proposed GPA-2
Decision Tree with Gini Index (DT-GI)	47.49%	91.51%	92.73%	55.54%	60.28%	57.92%
Decision Tree with Information Gain (DT-IG)	45.19%	94.24%	92.29%	54.72%	61.27%	59.27%
5-Nearest Neighbour (5-NN)	48.97%	92.78%	93.19%	58.21%	71.59%	69.46%
Naive Bayes	41.95%	74.44%	73.30%	73.60%	85.14%	81.88%
Neural Network	48.67%	91.53%	92.41%	57.09%	56.16%	53.04%
Random Forest with Gini Index (RF-GI)	48.19%	93.45%	92.89%	68.35%	72.83%	72.71%
Random Forest with Information Gain (RF-IG)	49.93%	95.49%	93.03%	68.74%	73.19%	75.20%

Here,  $TP$  denotes True Positive,  $FP$  denotes False Positive,  $TN$  denotes True Negative and  $FN$  denotes False Negative. Again,  $P_o$  is the observed proportionate agreement,  $P_e$  is the expected proportionate agreement. Equation (7) is used in Table 5 and equation (11) in Table 6 for multi-class performance measurement.

### 5.1. Objective 1 - predicting HSC examination performance using classifiers

To predict students' performance, we used different classifiers. Table 5 represents the classification performance for the board GPA and our proposed GPA for both datasets. From Table 5, we can see that in both of our proposed approaches accuracy is nearly twice that of the board-obtained GPA for the college dataset. Also in a synthetic dataset, our proposed GPA provides better accuracy than the board GPA. In the case of college dataset, for board GPA Random Forest with Gini Index (RF-GI) gave an accuracy of 57.28% which is higher than other classifiers. For proposed GPA-1, RF-GI has the highest accuracy which is 96.45%. For proposed GPA-2, Neural Networks have the highest accuracy which is 95.48%. Additionally, RF-GI performs better overall in both of our proposed methods. However, Neural Networks in the synthetic dataset offer the highest accuracy for both the proposed GPA-1 and the board GPA, at 86.5% and 79.6%, respectively. On the other side, Naive Bayes offers the maximum accuracy for the proposed GPA-2, which is 83.2%.

Table 6 represents another performance measuring tool called Cohen's kappa. From this table, we can see that for the college dataset, the highest kappa value for our proposed GPA-1 is obtained for Neural Networks which is 0.604. Also, for our second proposed GPA highest kappa

value is obtained for 5-Nearest Neighbour which is 0.593. Whereas the board GPA's highest kappa value is 0.210, obtained for RF-GI classifier. However, Neural Networks provide the highest kappa values for both the proposed GPA-1 and the board GPA in the synthetic dataset, at 0.681 and 0.719, respectively. On the other hand, Naive Bayes provides a maximum kappa value of 0.675 for the proposed GPA-2. Here, we can see that the board GPA's kappa value is lower than both of our proposed GPAs. Hence, it can also be said that our proposed GPA calculations are more closely related to the student's internal exam scores than their board GPA.

Table 7 represents F1-score for board GPA and proposed GPA for both of our datasets. From the analysis of the college dataset, we can see that, for board GPA and proposed GPA-1, Random Forest with Information Gain (RF-IG) offers the highest value. For the board GPA, it is 49.93% and for the proposed GPA-1 it is 95.49%. On the other hand, 5-Nearest Neighbour offers the highest F1-score for the proposed GPA-2 which is 93.19%. In contrast, Naive Bayes provides the highest F1-score for board GPA and two proposed GPA calculations, which are 73.6%, 85.14%, and 81.88%, respectively, in the synthetic dataset. Table 7 shows that the proposed GPA performs better overall in terms of F1-score. In brief, the analytical outcomes suggest that the proposed GPA calculations are more closely aligned with students' internal examination performance than with their board GPA. Tables 8, 9, and 10 represent the confusion matrix of the board GPA and our two proposed GPA calculations. Through the confusion matrices, we can identify the right and wrong predictions for each class made by the classifiers.

**Table 8**  
Confusion matrix for different classifiers using board GPA.

Classifier Names	Classifier's Prediction	True Poor	True Average	True Good	Class Precision	Class Recall
Decision Tree with Gini Index (DT-GI)	Predicted Poor	53	13	58	42.74%	45.30%
	Predicted Average	10	4	15	13.79%	10.00%
	Predicted Good	54	23	79	50.64%	51.90%
Decision Tree with Information Gain (DT-IG)	Predicted Poor	58	13	46	49.57%	49.57%
	Predicted Average	11	6	14	19.35%	15.00%
	Predicted Good	48	21	92	57.14%	60.53%
5-Nearest Neighbour (5-NN)	Predicted Poor	65	17	41	52.85%	55.56%
	Predicted Average	16	10	33	16.95%	25.00%
	Predicted Good	36	13	78	61.42%	51.32%
Naive Bayes	Predicted Poor	44	7	18	63.77%	37.61%
	Predicted Average	30	13	50	13.98%	32.50%
	Predicted Good	43	20	84	57.14%	55.26%
Neural Network	Predicted Poor	58	14	48	48.33%	49.57%
	Predicted Average	8	2	12	9.09%	5.00%
	Predicted Good	51	24	92	55.09%	60.53%
Random Forest with Gini Index (RF-GI)	Predicted Poor	52	11	27	57.78%	44.44%
	Predicted Average	0	0	0	0.00%	0.00%
	Predicted Good	65	29	125	57.08%	82.24%
Random Forest with Information Gain (RF-IG)	Predicted Poor	46	10	27	55.42%	39.32%
	Predicted Average	2	0	0	0.00%	0.00%
	Predicted Good	69	30	125	55.80%	82.24%

**Table 9**  
Confusion matrix for different classifiers using proposed GPA-1.

Classifier Names	Classifier's Prediction	True Poor	True Average	True Good	Class Precision	Class Recall
Decision Tree with Gini Index (DT-GI)	Predicted Poor	289	8	0	97.31%	98.97%
	Predicted Average	3	4	3	40.00%	30.77%
	Predicted Good	0	1	1	50.00%	25.00%
Decision Tree with Information Gain (DT-IG)	Predicted Poor	288	6	1	97.63%	98.63%
	Predicted Average	3	6	2	54.55%	46.15%
	Predicted Good	1	1	1	33.33%	25.00%
5-Nearest Neighbour (5-NN)	Predicted Poor	288	6	0	97.63%	98.63%
	Predicted Average	4	6	1	54.55%	46.15%
	Predicted Good	0	1	3	75.00%	75.00%
Naive Bayes	Predicted Poor	263	2	0	99.25%	90.07%
	Predicted Average	29	11	2	26.19%	84.62%
	Predicted Good	0	0	2	100.00%	50.00%
Neural Network	Predicted Poor	289	5	0	98.30%	98.97%
	Predicted Average	3	8	4	53.33%	61.54%
	Predicted Good	0	0	0	0.00%	0.00%
Random Forest with Gini Index (RF-GI)	Predicted Poor	292	8	0	97.33%	100.00%
	Predicted Average	0	5	3	62.50%	38.46%
	Predicted Good	0	0	1	100.00%	25.00%
Random Forest with Information Gain (RF-IG)	Predicted Poor	292	9	0	97.01%	100.00%
	Predicted Average	0	4	2	66.67%	30.77%
	Predicted Good	0	0	2	100.00%	50.00%

## 5.2. Objective 2 - relating predictor subjects and their impact with students' performance prediction

For the second research objective, we need to connect decision tree visualization with the prediction performance of board and proposed GPA calculations. After the distillation of decision tree (where attribute selection was performed by information gain, gini index, and accuracy) and merging decision tree graphs outcome (selective attributes) using voting method, an impact score is assigned per subject as its total node count shown in Table 11. This resembles how much each subject is important for student performance. The higher the score, the higher the subject's importance.

Table 11 summarizes the importance of every subject for student performance. After analysis, the most important subject for the board

GPA and proposed GPA-1 prediction is English and for the proposed GPA-2 is Bangla. Whereas the least important subject for board GPA is Optional, for proposed GPA-1 are Civics and Optional, for proposed GPA-2 are Civics and Islamic History. Besides, board GPA classification uses all of the 7 subjects with maximum total node counts 103, proposed GPA-1 classification uses 5 out of 7 subjects with only total node counts 12, and proposed GPA-2 classification uses 7 out of 7 subjects with only total node counts 17. Therefore, it is understandable that decision tree model has too many branches that lead to overfitting and poor generalization on unseen data (poor model accuracy) in board GPA classification. On the other hand, the proposed GPA calculations have efficient attribute selection and better model generalization as well. In a summary, students performance by proposed GPA calculations and subjects have stronger correlation than board GPA.

**Table 10**  
Confusion matrix for different classifiers using proposed GPA-2.

Classifier Names	Classifier's Prediction	True Poor	True Average	True Good	Class Precision	Class Recall
Decision Tree with Gini Index (DT-GI)	Predicted Poor	285	7	1	97.27%	97.94%
	Predicted Average	6	6	4	37.50%	46.15%
	Predicted Good	0	0	0	0.00%	0.00%
Decision Tree with Information Gain (DT-IG)	Predicted Poor	286	7	0	97.61%	98.28%
	Predicted Average	5	6	5	37.50%	46.15%
	Predicted Good	0	0	0	0.00%	0.00%
5-Nearest Neighbour (5-NN)	Predicted Poor	283	4	0	98.61%	97.25%
	Predicted Average	8	7	1	43.75%	53.85%
	Predicted Good	0	2	4	66.67%	80.00%
Naive Bayes	Predicted Poor	265	0	0	100.00%	91.07%
	Predicted Average	26	11	3	27.50%	84.62%
	Predicted Good	0	2	2	50.00%	40.00%
Neural Network	Predicted Poor	289	7	0	97.64%	99.31%
	Predicted Average	2	6	5	46.15%	46.15%
	Predicted Good	0	0	0	0.00%	0.00%
Random Forest with Gini Index (RF-GI)	Predicted Poor	290	9	0	96.99%	99.66%
	Predicted Average	1	4	5	40.00%	30.77%
	Predicted Good	0	0	0	0.00%	0.00%
Random Forest with Information Gain (RF-IG)	Predicted Poor	291	10	1	96.36%	100.00%
	Predicted Average	0	3	4	42.86%	23.08%
	Predicted Good	0	0	2	0.00%	0.00%

**Table 11**  
Predictor subjects impact score on prediction performance.

Subjects	Impact score on board GPA	Impact score on proposed GPA-1	Impact score on proposed GPA-2
English	23	6	3
Bangla	14	1	5
ICT	17	3	2
Islamic History	17	1	1
Sociology	12	1	3
Civics	12	0	1
Optional	8	0	2

In addition, from the Figs. 1, 2, 3, 4, and 5, it is seen that board GPA classification accuracy is below 60% using 27 out of 28 attributes. On the other hand, proposed GPA-1 and 2 have almost double classification accuracy (about 90%) using only 9 attributes. This indicates that proposed GPA calculations have significantly high correlation with students internal examination performance than board GPA. The proposed two GPA calculations relate more to student performance because they have no assumption about good results in JSC and SSC examinations, which indicates good result in HSC examination and based on board GPA (Alo, 2020).

### 5.3. Objective 3 - continuous performance progression of students

Third research objective investigates for valuable insights about group wise and individual students performance progression in the basis of subjects through internal academic examination. At first, we determined individual students' (random student with roll x) GPA booster, degrader, and at-risk subjects in half yearly, yearly, pretest, and test examinations as shown in Table 12. Besides, a visual perspective of

individual student performance in 7 subjects within each internal examinations is shown in Fig. 7 which reveals strong and weak points of individual student. Secondly, categorizing all subjects into performance booster, degraded, and at-risk subjects by number of overall affected students in all four internal examinations as shown in Fig. 8. Fig. 8a shows that English has affected no student as performance booster, 309 students as performance degrader, and for 304 students it was at-risk subject in half yearly examination. Similarly, Fig. 8b shows that Optional subject has affected 240 students as performance booster, 69 students as performance degrader, and for 65 students as at-risk subject in yearly examination. Fig. 8c presents pretest examination statistics, where it is clear that English has affected 3 students as performance booster, 306 students as performance degrader, and 295 students as at-risk subject. Similarly, Fig. 8d shows that English is the most performance degrader and at-risk subject in test examination as well.

Moreover, Table 13 shows student's respective half yearly, yearly, pretest, and test examination's current performance standing and performance booster subjects with minimum effort. In Table 13, roll y (random student) has poor performance currently after half yearly examination and for performance improvement to average, he/she must do well in Optional, Islamic History, Sociology, Civics, ICT, Bangla, and English. Besides, roll y has average performance currently after yearly examination and for performance improvement to be good, he/she must do well in Civics, Islamic History, and English. Again, roll y has poor performance currently after the pretest examination and for performance improvement to average, he/she must do well in Bangla and English. Similarly, roll y has poor performance currently after the test examination and for performance improvement to average, he/she must do well in English, Civics, and Bangla.

**Table 12**  
Individual student's booster, degrader, and at-risks subjects.

Student roll	Examination name	Booster subjects	Degrader subjects	At-risk subjects
x (Random student)	HalfYearly	Civics, Islamic History, Optional	English, Bangla, ICT, Sociology	English, Bangla, ICT, Sociology
	Yearly	Bangla, ICT, Sociology, Optional	English, Civics, Islamic History	English, Civics, Islamic History
	PreTest	Sociology, Islamic History, Optional	English, Bangla, ICT, Civics	English, Bangla, ICT, Civics
	Test	Bangla, Sociology, Islamic History, Optional	English, ICT, Civics	English, Bangla, ICT, Civics, Sociology, Optional

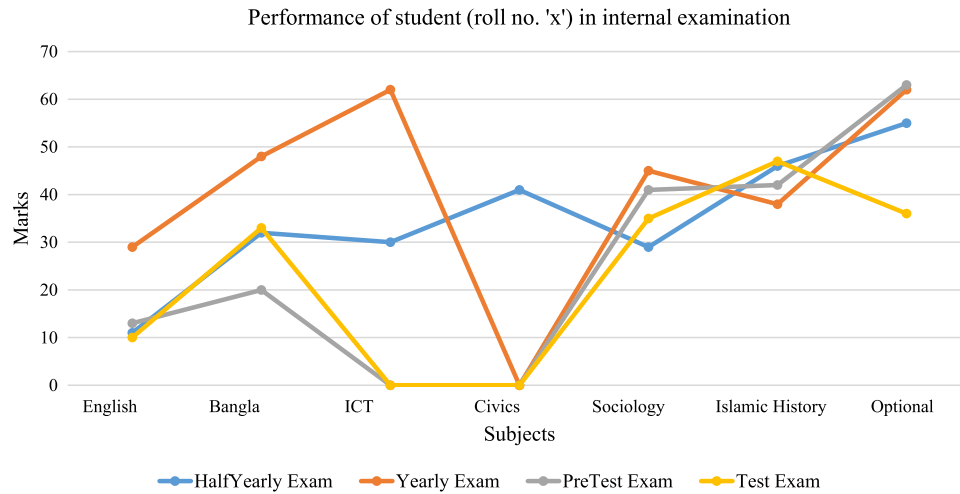


Fig. 7. Individual student's performance in all internal examinations.

Table 13

Individual students' performance improving and low effort subjects.

Student roll	Examination name	Current performance	Low effort subjects
y (Random student)	HalfYearly	Poor	Optional, Islamic History, Sociology, Civics, ICT, Bangla, English
	Yearly	Average	Civics, Islamic History, English
	PreTest	Poor	Bangla, English
	Test	Poor	English, Civics, Bangla

Table 14

Classification accuracy by classifiers of consistent performer students (earlier detection).

Classifier	Accuracy (Yearly GPA)	Accuracy (Pretest GPA)	Accuracy (Test GPA)
DT-GI	84.16%	89.66%	90.94%
DT-IG	83.50%	88.04%	90.61%
5-NN	84.79%	89.98%	90.30%
Naive Bayes	66.68%	80.58%	73.13%
Neural Network	86.41%	90.30%	91.91%
RF-GI	87.06%	90.95%	91.59%
RF-IG	86.42%	91.28%	91.60%

Table 15

Weighted-F1 score by classifiers of consistent performer students (earlier detection).

Classifier	Weighted-F1 (Yearly GPA)	Weighted-F1 (Pretest GPA)	Weighted-F1 (Test GPA)
DT-GI	82.18%	88.82%	86.46%
DT-IG	79.42%	85.44%	87.54%
5-NN	81.03%	88.30%	86.02%
Naive Bayes	74.34%	79.85%	83.03%
Neural Network	78.69%	84.75%	85.80%
RF-GI	83.94%	89.84%	88.29%
RF-IG	83.00%	89.81%	88.16%

#### 5.4. Objective 4 - consistent performance pattern early detection

Table 14 represents the classifier accuracy and Table 15 represents weighted F1-score. These two tables represent how precisely consistent performer students can be detected using their consistent performance in two consecutive previous internal examinations. From Table 14, it can be seen that Naive Bayes has the lowest classification accuracy performance while RF-GI has the highest classification accuracy performance which is absolutely satisfactory performance. Table 15 also supports selection of RF-GI model. Because it has the highest weighted-F1 score to correctly detect consistent performer students after consecutive internal examination periods (3 scenarios: after yearly examination, after pretest examination, after test examination). It has been found from the analysis of confusion matrix of RF-GI that there were 6 consistent good performer, 3 consistent average performer, and 260 consistent poor performer students after half yearly and yearly examination. This model has found consistent performance pattern of 269 students out of 309 students. Similarly, there were total of 281 consistent performers with 5 consistent good students, 10 consistent average students, and 266 consistent poor-performing students after yearly and pretest examination. In the same way, a total of 283 consistent performer student with 4 good, 2 average and 277 poor performer were there after pretest and test examination had occurred. This indicates that, after the yearly examination and before the pretest examination, about 87.05% of consistent

performer students were there. Through early intervention, 260 poor performers and 3 average performers could be handled for performance improvement. Also after the pretest examination and before the test examination, about 90.94% of consistent performer students were there. Thus, 266 poor performers and 10 average performers could be helped for performance improvement with early intervention. Finally, 91.58% early detected consistent performing students could be treated for performance boost between the period between test examination and HSC examination where 4 good, 2 average, and 277 consistent poor performer students might get a chance to do better result. Ultimately, the goal is to do better performance in HSC final examination. Consistent performance patterns appear to be a very important insight for the early identification of at-risk, struggling, and good-performing students. By appropriate guidance and monitoring, the performance of these students can be improved.

#### 6. Limitations and future work scopes

One of the limitations of this research is absence of students' socioeconomic and demographic features in the dataset. These data can greatly impact on academic performance of students. Besides dataset is quite imbalanced because it is collected from real world situations resulting in a few instances of good result-holder students. Also, no class imbalance and feature selection related study is addressed in this re-

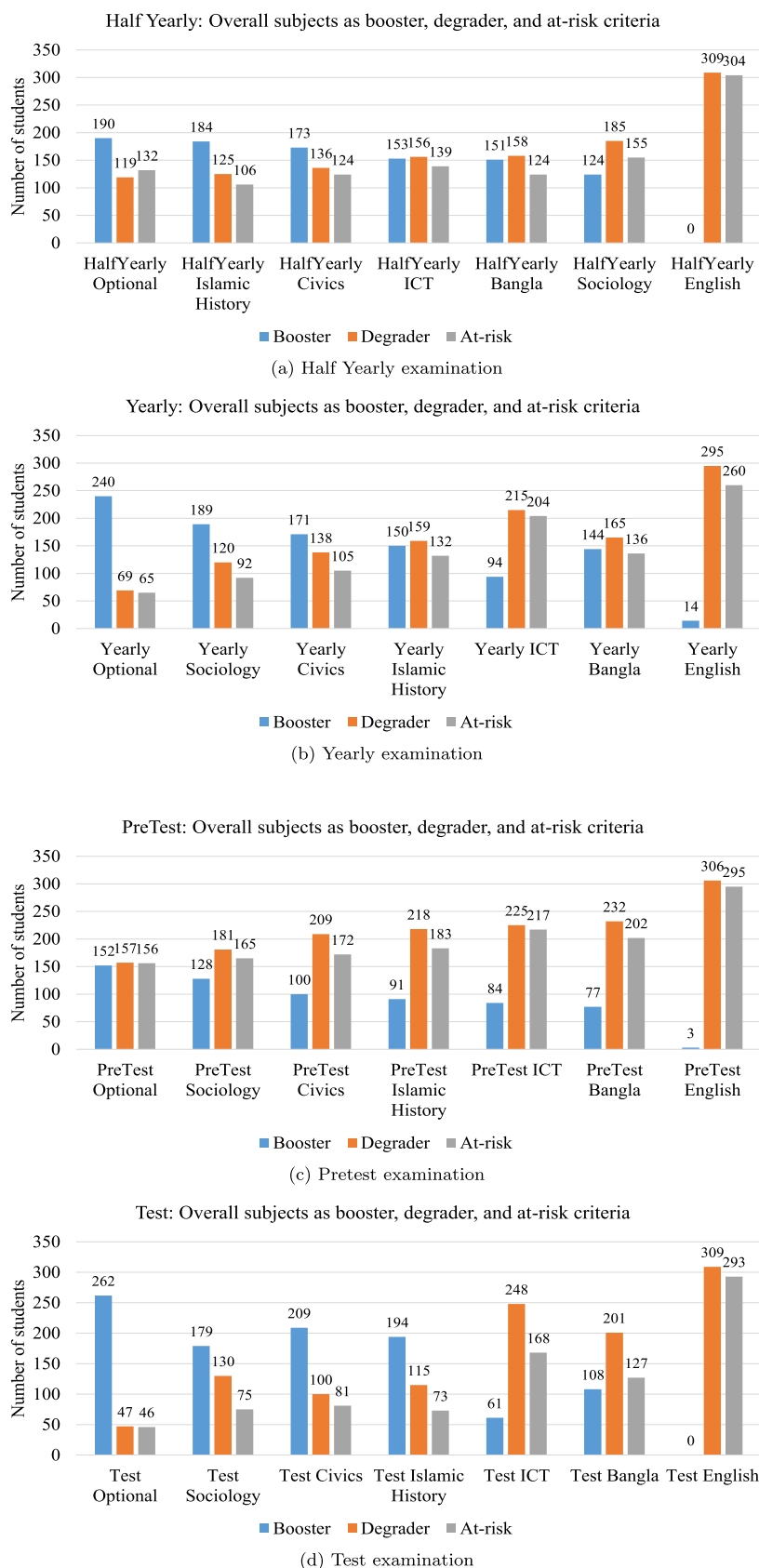


Fig. 8. Overall subject as booster, degrader, and at-risk criteria.

search. Moreover, one combined final mark per subject is considered here in place of subject-wise written, multiple-choice question (MCQ), practical marks to avoid analysis complexity. Increasing student data samples and collecting students' socioeconomic, and demographic features in the dataset using surveys can be a crucial future work scope. In addition, generalization of the findings for more college students' data can be another crucial work direction. Association rule mining during classification can be used to gain insights into the factors that contribute to student performance.

Besides proposing academic policies, initiatives through deep insights for performance improvement can be an extension of this work. Moreover, threshold-based approach for classification purposes and comparing among classification performance of traditional classifiers can be another work direction. Furthermore, ensemble or voting methods against traditional classification model comparison can be possible to implement. After all, the application of a novel approach to address typical performance progression can be accomplished. Finally, consistent performer student detection using previous examination performance trends can be an interesting working trajectory.

## 7. Theoretical and pedagogical implications of this study

The study conducted in this paper has both theoretical and pedagogical implications. The implications are stated below in accordance with the four research objectives of this study.

**Research Objective 1 (Machine learning models with several performance evaluation metrics):** We proposed two reasonable techniques for calculation of HSC GPA based on college internal examination performances. Thus, a comparison between proposed and board GPA can be conducted and relevancy with internal college performance can be found. This also helps to observe how a pandemic situation like COVID-19 had affected the students' results.

**Research Objective 2 (Visualization and interpretation of decision tree graphs):** Finding impact factors of subjects on student results can be helpful for both college authority and students to find out the strong and weak zones of subjects.

**Research Objective 3 (Extracting subject-wise performance knowledge by using novel approach of proposed python model):** Finding the performance progression of students after each internal examination can surely guide upcoming performance boost. Outcomes like categorization of each subject as performance booster, degrader or at-risk subject, and low effort performance booster subjects are crucial aspects of student performance that can be utilized by college authorities to act wisely for performance improvement.

**Research Objective 4 (Consistent performance trends detection using machine learning models and interpretation of confusion matrix):** Identifying groups of students who perform consistently similar after consecutive internal examinations is an interesting performance aspect to distinguish students as consistently good, consistently average, or at-risk of failure. Therefore, college authorities can surely monitor, motivate, and provide special care to each group of students by using these valuable insights.

## 8. Conclusion

This study investigated four research objectives to provide valuable insights about academic performance. This might help to improve the student's academic performance and institutional educational quality as well. The first objective concerns predicting students' HSC examination performance using internal examination marks only, while no socio-economic data are available. The classification results based on both college and synthetic dataset show that the proposed GPA calculations have better classification performance than board GPA. In other words, the proposed GPA is more relevant to internal examination performance than the board GPA. Besides, the second objective aims to identify the impact of subjects on student performance. Visualization

of decision tree shows that proposed GPA has better attribute selection and better classification than board GPA. Board GPA has over-fitted decision tree and poor attribute selection with poor classification performance. Impact of subjects on classification performance shows that the most important subjects for board GPA are English, ICT, and Islamic history, for proposed GPA-1 are English, ICT, and Bangla, and for proposed GPA-2 are Bangla, English, and Sociology. This resembles the importance of subjects for GPA classification. The third objective investigates how every student's subject-wise academic performance progresses over the two-year degree. Our proposed method for continuous performance progression can identify every student's performance booster, degraded, and at-risk subjects. It also represents individual student's low-effort performance booster subjects. Besides, every subject can be categorized as booster, degrader, and at-risk subject based on the number of affected students. These outcomes can certainly aid in comprehending the subject and current student performance relationship. Therefore, a strategy to help struggling students can be deduced based on valuable informative results. Through more efficient academic assessment methods and improved teaching strategies, teachers will be able to more readily identify children who are at-risk and take the required measurements to support their academic performance recovery. The fourth objective examines consistent performance patterns based on two consecutive internal examination performances. Classification results indicate that RF-GI has the highest classification performance and has correctly detected the highest amount of consistent performer students. According to confusion matrix analysis, there are 86.0% consistent performer students after yearly examination whereas 90.9% after pretest and 91.6% consistent performer students after test examination. Earlier detection of consistent performer students can significantly help to identify bright, struggling, and at-risk students. Consequently, college authorities can formulate academic strategies to guide, monitor and motivate bright students as well as take immediate steps for performance improvement of struggling and at-risk students.

## Statements on open data, ethics and conflict of interest

The study was approved by an ethical committee with ID: EIIN 1\*27\*9. Informed consent was obtained from all participants, and their privacy rights were strictly observed.

## List of acronyms

Acronyms	Definition
JSC	Junior School Certificate
SSC	Secondary School Certificate
HSC	Higher Secondary Certificate
EDM	Educational Data Mining
GPA	Grade Point Average
DT	Decision Tree
RF	Random Forest
NN	Neural Network
KNN	K-Nearest Neighbor
IG	Information Gain
GI	Gini Index

## CRediT authorship contribution statement

**Sazol Sarker:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mahit Kumar Paul:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Sheikh Tasnimul Hasan Thasin:** Writing – original draft, Visualization, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Md. Al Mehedi Hasan:** Writing – review & editing, Validation, Supervision, Formal analysis.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Alo, P. (2020). Hsc results of 2020 in January after issuance of ordinance: Dipu moni. *Prothom Alo*. Retrieved from <https://en.prothomalo.com/youth/education/hsc-results-of-2020-in-january-after-issuance-of-ordinance-dipu-moni>.
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>.
- Baker, R. (2010). Data mining. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (third edition) (pp. 112–118). Oxford: Elsevier.
- Baradwaj, B. K., & Pal, S. (2011). Mining educational data to analyze students performance. *International Journal of Advanced Computer Science and Applications*, 2. <https://doi.org/10.14569/IJACSA.2011.020609>.
- Beale, R., & Jackson, T. (1990). *Neural computing-an introduction*. CRC Press.
- Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass prediction model for student grade prediction using machine learning. *IEEE Access*, 9, 95608–95621.
- Czibula, G., Mihai, A., & Crivei, L. M. (2019). S prar: A novel relational association rule mining classification model applied for academic performance prediction. *Procedia Computer Science*, 159, 20–29.
- Duda, R. O., Hart, P. E., et al. (1973). *Pattern classification and scene analysis*, Vol. 3. Wiley New York.
- Edu, T. L. (2023). Importance of education. Retrieved from <https://leverageedu.com/blog/importance-of-education/>.
- Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, 340, 1177–1178. <https://doi.org/10.1126/science.1236536>.
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29, 131–163. <https://doi.org/10.1023/A:1007465528199>.
- Golding, P., & Donaldson, O. (2006). Predicting academic performance. In *Proceedings. Frontiers in education. 36th annual conference* (pp. 21–26). IEEE.
- Guzmán, E., Conejo, R., & Pérez-de-la Cruz, J.-L. (2007). Improving student performance using self-assessment tests. *IEEE Intelligent Systems*, 22, 46–52. <https://doi.org/10.1109/MIS.2007.71>.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques*. Morgan Kaufmann.
- Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers and Education*, 61, 133–145. <https://doi.org/10.1016/j.compedu.2012.08.015>.
- Kaunang, F. J., & Rotikan, R. (2018). Students' academic performance prediction using data mining. In *2018 third international conference on informatics and computing (ICIC)* (pp. 1–5). IEEE.
- Kono, H., Sawada, Y., & Shonchoy, A. S. (2018). Primary, secondary, and tertiary education in Bangladesh: Achievements and challenges. In *Economic and social development of Bangladesh: Miracle and challenges* (pp. 135–149).
- Kramer, O., & Kramer, O. (2013). K-nearest neighbors. In *Dimensionality reduction with unsupervised nearest neighbors* (pp. 13–23).
- Leung, R. K., Wang, Y., Ma, R. C., Luk, A. O., Lam, V., Ng, M., So, W. Y., Tsui, S. K., & Chan, J. C. (2013). Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: A prospective case-control cohort analysis. *BMC Nephrology*, 14, 1–9. <https://doi.org/10.1186/1471-2369-14-162>.
- Liu, H., Zhou, M., Lu, X. S., & Yao, C. (2018). Weighted Gini index feature selection method for imbalanced data. In *2018 IEEE 15th international conference on networking, sensing and control (ICNSC)* (pp. 1–6). IEEE.
- Mashiloane, L., & Mchunu, M. (2013). Mining for marks: A comparison of classification algorithms when predicting academic performance to identify "students at risk". In *Mining intelligence and knowledge exploration: First international conference, MIKE 2013, Tamil Nadu, India, December 18–20, 2013. Proceedings* (pp. 541–552). Springer.
- Maxwell, J. A. (2012). The importance of qualitative research for causal explanation in education. *Qualitative Inquiry*, 18, 655–661. <https://doi.org/10.1177/10778004124528>.
- Meghji, A. F., Mahoto, N. A., Asiri, Y., Alshahrani, H., Sulaiman, A., & Shaikh, A. (2023). Early detection of student degree-level academic performance using educational data mining. *PeerJ Computer Science*, 9, Article e1294.
- Mishra, T., Kumar, D., & Gupta, S. (2014). Mining students' data for prediction performance. In *2014 fourth international conference on advanced computing & communication technologies* (pp. 255–262). IEEE.
- Müller, B., Reinhardt, J., & Strickland, M. T. (1995). *Neural networks: An introduction*. Springer Science & Business Media.
- Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the Gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41, 77–93. <https://doi.org/10.1023/B:AMAI.0000018580.96245.c6>.
- Ren, J., Lee, S. D., Chen, X., Kao, B., Cheng, R., & Cheung, D. (2009). Naive Bayes classification of uncertain data. In *2009 ninth IEEE international conference on data mining* (pp. 944–949). IEEE.
- Star, T. D. (2021). Hsc equivalent exams: Govt can now publish results without exam. *The Daily Star*. Retrieved from <https://www.thedailystar.net/frontpage/news/hsc-equivalent-exams-govt-can-now-publish-results-without-exam-2033245>.
- Sun, S., & Huang, R. (2010). An adaptive k-nearest neighbor algorithm. In *2010 seventh international conference on fuzzy systems and knowledge discovery, volume 1* (pp. 91–94). IEEE.
- Tangirala, S. (2020). Evaluating the impact of Gini index and information gain on classification using decision tree classifier algorithm\*. *International Journal of Advanced Computer Science and Applications*, 11. Retrieved from <https://api.semanticscholar.org/CorpusID:212657021>.
- Tasnim, N., Paul, M. K., & Sattar, A. H. M. S. (2019). Identification of drop out students using educational data mining. In *2019 international conference on electrical, computer and communication engineering (ECCE)* (pp. 1–5).
- The Finance Today (2023). Public varsities caught in further session jam. Retrieved from <https://www.thefinancetoday.net/article/education/15811/Public-varsities-caught-in-further-session-jam>. (Accessed 21 April 2024).
- Tribune, D. (2020). No hsc equivalent exams this year: Students to. *Dhaka Tribune*. Retrieved from <https://www.dhakatribune.com/bangladesh/education/225752/no-hsc-equivalent-exams-this-year-students-to>.
- World Health Organization (2019). Coronavirus. *WHO website*. Retrieved from [https://www.who.int/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/health-topics/coronavirus#tab=tab_1). (Accessed 18 May 2023).
- Yang, F., & Li, F. W. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers and Education*, 123, 97–108. <https://doi.org/10.1016/j.compedu.2018.04.006>.
- Yuan, P., Chen, Y., Jin, H., & Huang, L. (2008). Msvm-knn: Combining svm and k-nn for multi-class text classification. In *IEEE international workshop on semantic computing and systems* (pp. 133–140). IEEE.