Predicting Undergraduate Admission: A Case Study in Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh

Md. Protikuzzaman[1], Mrinal Kanti Baowaly[2], Maloy Kumar Devnath[3], Bikash Chandra Singh[4]

Department of Computer Science and Engineering
Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh[1, 2, 3]
Department of Information and Communication Technology, Islamic University, Kushtia, Bangladesh[4]

Abstract—The university admission tests find the applicant's ability to admit to the desired university. Nowadays, there is a huge competition in the university admission tests. The failure in the admission tests makes an examinee depressed. This paper proposes a method that predicts undergraduate admission in universities. It can help students to improve their preparation to get a chance at their desired university. Many factors are responsible for the failure or success in an admission test. Educational data mining helps us to analyze and extract information from these factors. Here, the authors apply three machine learning algorithms XGBoost, LightGBM, and GBM on a collected dataset to estimate the probability of getting admission to the university after attending or before attending the admission test. They also evaluate and compare the performance levels of these three algorithms based on two different evaluation metrics – accuracy and F1 score. Furthermore, the authors explore the important factors which influence predicting undergraduate admission.

Keywords—Undergraduate admission; educational data mining; XGBoost; Light GBM; GBM; evaluation metrics

I. Introduction

In any country, an undergraduate admission test is one of the most important tests for the students. Students remain conscious about taking admission to their desired universities. In Bangladesh, students who passed the Higher Secondary Certificate (HSC) examination contest the undergraduate admission test. According to the year 2017, 8.01 lakh examinees passed the HSC exam [1] and competed to get admission in different public universities. Universities have their admission requirements for this purpose which are generally based on the students' grade point average (GPA) of the Secondary School Certificate (SSC) and HSC examination, GPA of various courses, etc. However, the total seats in public universities are not sufficient. According to the Ministry of Education of Bangladesh [1], the number of seats in the country's 37 public universities is around 60,000. As a result, about 7 lakhs 40 thousand students did not get the opportunity to study in public universities last year. Even those students who can apply and sit for the admission test do not have the guarantee of admission opportunities in the university because of the limited number of seats. Students have to overcome the barrier of admission test and qualify in the examination to secure their seats. Such students have to go through a long time of mental stress or illness before or after

the admission test. The authors realize that this issue cannot be completely removed. But with the aid of modern technologies and strategies e.g. educational data mining, this study can reduce the problem and make students aware of it early in the admission test. If any student can know the pre-examination and post-examination status of a particular university for undergraduate admission, it will be a great benefit for him/her to take the necessary steps to improve the admission test's performance so that he/she can get a chance at the desired university. The authors want to help the students to judge and improve themselves before or after the admission test using this system.

In this study, the authors use the concepts and techniques of data mining which is discovered useful and meaningful information from large-scale data collections [2, 3, 4]. Because of the growing data volume of educational knowledge, educational data mining has a rich area of application [5]. This research is conducted to measure the admission opportunity of a student in Bangabandhu Sheikh Mujibur Rahman Science and Technology University (BSMRSTU), Bangladesh. It is more authentically based on a real dataset collected from the engineering and science faculty students of BSMRSTU. Discovering knowledge from real data gives us a solution that helps students to improve their performance to get admission to BSMRSTU. The authors apply different data mining techniques [6] for a fruitful solution. Here, a total of 500 students' data is collected for this investigation. Though this research focuses on admission chance as the case study, the proposed approach is not restricted to it only. Moreover, this study extensively investigates all possible features or factors of an undergrad candidate and evaluates their impact for predicting admission. The main contributions of this thesis are:

- Developing an admission prediction system for the undergrad students in the engineering faculty at BSMRSTU, Bangladesh.

- Predicting the admission opportunity both before and after the admission test.

- Analyzing and evaluating the possible factors of an admission candidate that affect the admission chance.

The remaining sections of the paper are organized here. In Section II, related works are described. Section III introduces

our proposed model for the prediction. The experiment and results of the proposed technique have been described and presented in Section IV. Discussion of the results has been given in Section V. This paper includes the conclusion in Section VI.

II. RELATED WORKS

Researchers are working towards the modernization of the education system using education data mining technology. There is a survey [7] paper that depicts the most relevant studies using educational data mining. The researchers concentrate on the field of educational data mining as recent studies show that it is used for analyzing students' performance [8, 9, 10]. There are a number of past studies that focused on predicting admission in colleges or universities. A brief literature review of those studies is presented as follows.

Binu et al. [11] proposed a cloud-based data analysis and prediction system for predicting university admission. There were two modules in the proposed framework, i.e. A Hadoop MapReduce data storage module and an Artificial Neural Network to predict the chances. The data collected had attributes such as status, rank, board, quota, etc. The system did not use academic qualifications in the forecasting process. The neural network had two input nodes, one hidden layer with two nodes, and one output layer with two nodes.

Acharya et al. [12] proposed a comparative approach to predicting graduate admissions by developing four models of machine learning regression: linear regression, vector support, decision tree, and random forest. Roa et al. [13] built a College Admission Predictor in the form of a web application, taking as input the scores obtained by the applicant and his/her personal information and predicting potential college admissions as output.

Ghai [14] developed an American Graduate Admission Prediction model that allows students to choose an apt university by predicting whether or not they will be admitted to the university. Gupta et al. [15] developed a machine learning decision support system for the prediction of graduate admissions in the USA by taking account of certain parameters, including standardized tests, GPA, and Institute Reputation.

Mane and Ghorpade [16] designed a framework for predicting student admission to a particular college using a hybrid combination of Association Rule Mining and Pattern Growth Approach. Data source attributes included student details such as name, gender, caste, address, 10th mark, 12th mark, the score of Common Entry Test, name of pre-college, name of admitted college, and branch. Once valid association rules have been established, the prediction shall be made by the constraint of consequence during the generation of association rules.

Raut and Nichat [17] worked to predict students' performance based on a standard classification methodology, the Decision Tree. This method proposed a model where students take an online test and get an immediate answer (Pass / Fail) coupled with poor principles. The generalization of the sequential pattern mining algorithm was used for the evaluation of output. The decision tree developed by C4.5 is

used to assess the success of students and to identify them based on their marks. The author noted that this data mining research could help administrators find poor students and offer extra guidance before the final exam.

Arsad and Buniyamin [18] used the ANN model to forecast the academic success of Bachelor of Technology graduates. The research considered Grade Point (GP) of main subjects that students rate as inputs without taking into consideration their socioeconomic context, thus considering Grade Point Average (GPA) as production. Neural Network (NN) trained engineering graduate students GP to achieve targeted performance. This work showed that core themes have a significant impact on the final CGPA graduation.

Erdogan and Timor [19] used cluster analysis and k-means algorithm techniques to uncover the connection between student entrance test outcomes and their performance. Ktona et al. [20] used the mining association rule as one of the mining partnership tools to classify variables that influence the information acquired by high school students in the ITC course.

Devasia et al. [21] introduced an analysis to predict the success of students in the upcoming academic history test. Build a Web-based program. Nineteen of 700 student characteristics are used as input. When the marks of the pupil were entered, it was contrasted with the scores of the current student, and the ranking of the Naïve Bayes was used to determine the final score. It is noted that the qualification of mother and family income is strongly associated with student success. The collection of data sources, the detection of performance-influencing variables, the construction of a predictive model, and the testing of the model were proposed in the creation of an academic prediction model. The authors noted that this model should help minimize the ratio of loss and help to take appropriate steps against poor performance.

Ruby and David [22] developed a prediction model focused on the Multi-Layer Perceptron algorithm. Datasets were composed of 165 scientific, personal, and economic documents. The overall performance reached for all attributes was 52% and the chosen attributes were 33%.

Aziz et al. [23] created a prediction model that predicts the performance of the first-year computer science students. They used the Naïve Bayes classifier to build their prediction model. By using Naïve Bayes Classifier, it would predict the students' performance level as a categorical value; Poor, Average, and Good. The authors showed that the students' family income, gender, and hometown parameter were the important factors for students' academic performance.

Anuradha and Velmurugan [24] built a new method for predicting the students' final exam results. They applied statistical classification techniques. The experiment shows classifier Naïve Bayes performs better than the other classifiers. The author noted that data mining would improve student status and success at the educational institution.

Kaur et al. [25] used a classification algorithm to classify and viewed slow learners among students using predictive data mining models. From comprehensive literature reviews, variables that affect student success are identified. Both

parameters were used as input variables. Five classification algorithms MLP, Naïve Bayes, SMO, J48, and Reptree were applied to the datasets of high school students. MLP was found to have outperformed other classifiers at 75% accuracy. The authors showed that the students who had a computer and internet at home did better during the tests.

In this article, the authors want to determine the chance of a student gets admitted to a university: case study BSMRSTU, Bangladesh. For this, they are using three different machine learning methods that are most effective than others are described. This investigation is different from all the works mentioned above because this study investigates the examinees' possibility of getting a chance in a university before the admission test and after the admission test.

III. PROPOSED MODEL

The proposed system overview is displayed in Fig.1. It gives a summary of the possible model of the admission prediction. First, the authors collect data from Bangabandhu Sheikh Mujibur Rahman Science & Technology University (BSMRSTU). After collecting the data, they pre-process the data, extract the features. Then they apply supervised machine learning methods that train, validate the data, and extract knowledge from it. This study predicts the examinee's admission chance before and after the admission to the engineering faculty at BSMRSTU. Each part of this proposed system is described in the following subsections.

A. Data Collection

Having a university admission relies not only on the exam result of the students but also on many other considerations related to their social, economic, cultural, or geographical factors. First, the authors deeply analyze and recognize the causes that are liable or have an impact on admission. To collect data, it is prepared a set of 27 questionnaires as shown in the following subsection *B*. Then the authors provide these questionnaires to the students of BSMRSTU's various departments such as Computer Science and Engineering, Electrical and Electronic Engineering, Electronics and Telecommunication Engineering, Applied Chemistry and Chemical Engineering, Mathematics, Statistics, Chemistry, and Environmental Science & Disaster Management. The first four departments' data is considered for those who got admission in the Engineering faculty (also called A Unit) and the rest four departments' data for those who did not get admission in the Engineering faculty. The total collection is 500 students' data.
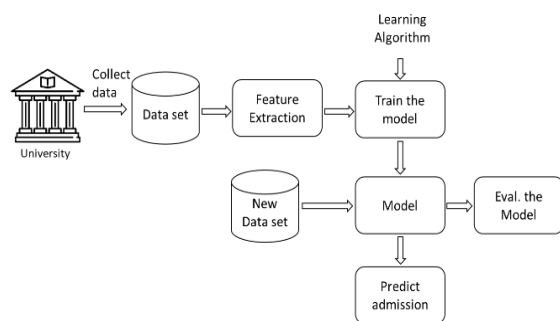

Fig. 1.   System Overview.

B. Questionnaires

1. What was your S.S.C GPA (Out of 5.00)?
2. What was your H.S.C GPA (Out of 5.00)?
3. What was your Physics Grade (H.S.C)?
4. What was your Chemistry Grade (H.S.C)?
5. What was your Math Grade (H.S.C)?
6. What was your English Grade (H.S.C)?
7. What was your college?
8. Where did your family live?
    a. Village
    b. Town
9. Name of the Village/Town?
10. District name?
11. About your family education?
    a. Highly Educated
    b. Educated
    c. Less Educated
12. About your family status?
    a. Rich
    b. Middle-Income
    c. Poor
13. Did you live with family at admission time?
    a. Yes
    b. No
14. Where did you live in admission time?
    a. Village
    b. Town
15. Name of the village/town name you lived during admission?
16. District name during admission?
17. Did you get any instruction from any coaching center during admission?
    a. Yes
    b. No
18. Did you have an inspector or motivator?
    a. Yes
    b. No
19. Did you involve in any political party?
    a. Yes
    b. No
20. Did you have any addiction to smoking or any other drug?
    a. Yes
    b. No
21. Did you have an internet facility during admission?
    a. Yes
    b. No
22. What was your daily wasting time (average) on social media or gameplay during admission time (in hours)?
23. What was your daily study time (average) for the admission test (in hours)?
24. Admission test year?
25. Second timer?
    a. Yes
    b. No
26. Expected scores/marks in A Unit-
27. Did you get admission in Unit A?
    a. Yes
    b. No

These features are grouped into two main categories without the target factor. These are (1) before engineering faculty admission (obtained marks in admission test is not included) (2) after engineering faculty admission (obtained marks in admission test is included.

C. Data Preprocessing

The authors prepared the collected data in tabular form from the questionnaire feedback of the students who participated during the data collection. They also applied some data cleaning techniques e.g. handling noise, outliers, missing values, and duplicate data to transform the raw data in a useful and efficient format. The authors considered each questionnaire as a distinct variable or feature for the dataset as shown in Table I. The authors split these 27 features into two categories before and after the admission test takes place to the BSMRSTU Engineering Faculty (Unit A). They allocate the first 25 variables as input features and the last factor (27th variable) as the output label to predict the chance of admission in the engineering faculty before the test. We then allocated the first 26 factors as input features and the last factor (27th variable) as the output label to predict the chance of admission in the engineering faculty after the test happened. The difference between these two categories is of one feature i.e. 'expected scores/marks in A Unit' (26th variable in Table I).

For identifying the most relevant input variables (feature selection) to predict undergraduate admission, the authors used the embedded methods which used built-in feature selection methods in machine learning algorithms. All three tree-based machine learning algorithms XGBoost, LightGBM, and GBM used in this investigation have their feature selection method.

D. Description of the Features

A description of the extracted features is given below. Some of the closely related features are discussed together.

S.S.C GPA: S.S.C result is an important factor for identifying a student's quality. It reflects the basic science knowledge of a student. To attend any university admission exam a student should have passed it successfully. So, this factor is important for predicting admission test results. The numeric grade is recorded out of 5.00.

H.S.C GPA: H.S.C result is the most important factor for identifying a student's quality. It also bears the science knowledge of a student. So this attribute is more important for predicting admission test results. The numeric grade is recorded out of 5.00.

Physics grade: A student's physics grade is the reflection of knowledge on the physics subject. In the BSMRSTU admission test, there are 30 marks on physics for the A unit. So, the grade of physics is more important for getting a chance in the engineering faculty of BSMRSTU. The numeric grade is recorded out of 5.00.

Chemistry grade: Chemistry grade is the reflection of knowledge on the chemistry of a student. In the BSMRSTU admission test of A unit, there are 20 marks on chemistry. So, the grade of chemistry is also important for getting a chance in

the engineering faculty of BSMRSTU. The numeric grade is taken out of 5.00.

Math grade: Math grade is the reflection of the knowledge of a student on mathematics. In the admission test of the BSMRSTU A unit, there are 30 marks on math. So, the grade of mathematics plays an important role in getting a chance in the engineering faculty of BSMRSTU. The numeric grade is taken out of 5.00.

English grade: English grade is the reflection of knowledge on the English of a student. We take the numeric grade out of 5.00.

TABLE I.　FEATURES LIST

	Feature No.	Features	Value
Before Exam (25 Input Features)	1.	S.S.C GPA	Out of 5.00
	2.	H.S.C GPA	Out of 5.00
	3.	H.S.C Physics Grade	Out of 5.00
	4.	H.S.C Chemistry Grade	Out of 5.00
	5.	H.S.C Math Grade	Out of 5.00
	6.	H.S.C English Grade	Out of 5.00
	7.	College Name	Distinct name
	8.	Family Living Position	Village or Town
	9.	Village/Town name	Distinct name
	10.	District name	Distinct name
	11.	Family Education	3 types
	12.	Family Status	3 types
	13.	Live With Family	Yes or No
	14.	Admission Time Live With Family	Yes or No
	15.	Village/Town Name During Admission	Distinct name
	16.	District Name During Admission	Distinct name
	17.	Coaching Center	Yes or No
	18.	Motivator	Yes or No
	19.	Political Party	Yes or No
	20.	Smoking/Drug	Yes or No
	21.	Internet Facility	Yes or No
	22.	Wasted Time	In Hour
	23.	Reading Time	In Hour
	24.	Admission Test Year	Year
	25.	Second Timer	Yes or No
After Exam(26 Features)	26.	Expected Scores/Marks in Engineering Faculty	Distinct marks
Target Factor	27.	Get admission in Engineering Faculty (Unit A)	Yes or No

College name: The college name is considered to know which college students are willing to admit to the engineering faculty of BSMRSTU. By this feature, we also know the college name that affected most in engineering faculty.

Living area: Bangladesh is a developing country. The development of the country is not equally distributed. The students living in the town area are more concerned about their future rather than the rural area students. Their parents are more conscious of their children's education. The students who live in town have more facilities rather than rural students. This is why it is divided this category into two sectors, living in town and living in the village. Also, including the village or town name and district name which students belong to that town or village.

Family education status: Family education is a great factor for removing the darkness of the mind of a child. Educated parents consider the education of their children as one of the basic needs of life and try hard and soul to provide them with it. They are so much concerned about their children's future. They want to see their children at least as a student in a public university. There are lots of parents who are not educated but they are conscious of their children's future. They also perform a partly role like an educated parent. Family education status also affects admission chances. In a highly educated family parents are more concerned about their child's education at the early stage. So, the authors divide this category into three sectors- Highly educated, educated, less educated.

Living status during admission: Each student during the admission test either lives in a mess or with a family. The students who live in the mess may have some problems to study. They need to maintain some rules of the mess which kills the time to study. On the other hand, the students who live with family have the extra facility to study and their parents may always take care of them. This is why we divide this attribute into two sectors yes and no. If yes that means, he/she lives with his/her family during admission. If no, then clearly mention the area and district name during admission.

Instruction center: This factor means where a student takes instruction for the admission test. In our country, there are three types of instruction centers to get admissions such as coaching center, batch, and private tuition. In the coaching center, there are lots of students who attend a class. So, the teacher cannot give a student special focus. But in batch or private tuition students can come closer to the teacher and the teacher can focus on each of the student's preparation. So for the admission test, this factor has special significance. We divide this factor into two categories take coaching or not.

Motivator: The motivator feature plays a vital role in getting a chance in the admission test. Students who do not have any motivator cannot understand what is needed for getting a chance at the university. Even most of them did not know about the university until they admit into a coaching center or batch. So motivator has great importance.

Political involvement: Students who are involved with the political party cannot be able to pay more attention to study because they remain busy with political meetings and processions. We divide this factor into two categories, yes or no.

Frustration and drug addiction: Frustration on anything especially on any matter and drug addiction can hamper a student's preparation. It also kills time to study. It can reduce the confidence of a student which is must be needed for getting a chance at the university. We divide this factor into two categories, yes or no.

Internet facility: Nowadays the internet is the most important thing or we can say it, teacher, for all types of students in all sectors. So, having an internet connection for collecting previous questions, learning difficult topics and many other purposes is important for a student who wants to get admitted to the university. For this reason, we divide this factor into two categories, yes or no.

Wasted time: Here wasted time means the time a student wastes on social media or playing online games. From many types of research, we know social media has a bad impact on students which hampers the study of a student especially for students who are taking admission tests, it can be like a curse because the students get a small amount of time for admission preparation. So, for this reason, we consider this factor to getting a chance in the admission test. We take this factor individually.

Study time: Study time is the main factor for getting chance in admission test because who studies more time have more opportunity to learn more things. We take this factor individually.

Admission test year: We can see that some years' questions are easy for getting a chance and some years' questions are comparatively hard. So the admission test year includes in which year students take part in the admission test.

Second timer: Students who face the admission test the second time have more experience and get more time for taking preparation than the students who face the admission test the first time. So we divide this factor into two categories, yes or no.

Admission test score: An admission test score is most important to predict the examinee's admission after the exam but before the exam, it is not necessary.

Admission test result: This is the main and most important factor which is also called the target factor. Our neural network will predict if a student will get a chance or not. We divide this factor into two categories, yes or no.

IV. Experiment and Results

Since the experimental outcome is very significant in any type of research, all the researchers want to achieve the highest level of accuracy according to their work. This level of accuracy can be different for using different algorithms and methodologies. Researchers must select the algorithm and approach that will provide the highest level of accuracy for the relevant study. In this investigation, the authors predict the admission of the examinees by using different types of supervised learning algorithms. This study applies some advanced algorithms, i.e., XGBoost, LightGBM, Gradient

Boosting Machine to train and validate the predictive model. The authors used k-fold cross-validation (k=5) for validating the model. Therefore, error percentages are lower.

Throughout this experiment, the authors use two specific different measures or metrics to evaluate the quality of classification: accuracy and F1-score based on the following Equations (1) and (4), respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

A. Classification Results before the Exam

This study used three machine learning techniques mentioned above to predict the possibility of getting admission in the engineering faculty at BSMRSTU before participating in the admission test. In this case, the obtained marks feature was not in the dataset because this investigation was before the test. The evaluation results to predict admission on test data are summarised in Table II. Note that the accuracy and F1 score are not high, the maximum value is nearly 60 in this case. It is justified as this model is trained and evaluated before the admission test and the authors do not have the admission test score. Nevertheless, the applicant can assess himself to some extent using this model before the admission test.

B. Classification Results after the Exam

This investigation then performs to predict admission opportunities in the engineering faculty at BSMRSTU after participating in the admission test. In this case, expected obtained marks in the exam is used because it is now known to the applicants. The evaluation results are given in Table III. GBM model achieves the highest score 95%. It means the proposed model using the GBM algorithm can accurately predict the admission chance of the applicants after the admission test.

C. Feature Importance before the Exam

The features' importance of the proposed model can be found by using the feature importance property of the model. Feature importance gives a score for each feature of our data, the higher the score is more important or relevant to our output variable. Fig. 2 shows the important features for predicting the admission opportunity of the applicants before the test using the XGBoost learning algorithm. Fig. 3 shows the important features for predicting the admission opportunity of the applicants before the test using the LightGBM learning algorithm. Fig. 4 shows the important features for predicting the admission opportunity of the applicants before the test using the GBM learning algorithm.

The authors combined the feature importance of the aforementioned three learning algorithms. This study did the average feature importance (score) of each feature and plotted in Fig. 5. It shows the features that affect most for getting an

opportunity of admitting in engineering faculty before participating in the admission test and these are as follows:

- College name
- Town/Village name
- District name
- H.S.C GPA
- Reading time

TABLE II.   CLASSIFICATION RESULTS BEFORE THE ADMISSION TEST

Evaluation Measures	XGBoost	LightGBM	GBM
Accuracy	0.58	0.53	0.50
F1 score	0.56	0.52	0.47

TABLE III.   CLASSIFICATION RESULTS AFTER THE ADMISSION TEST

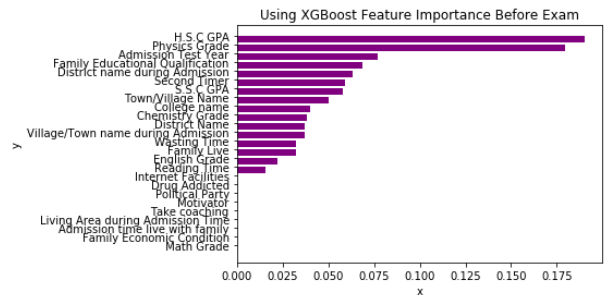| Evaluation Measures	XGBoost	LightGBM	GBM
Accuracy	0.87	0.93	0.95
F1 score	0.84	0.93	0.95

Fig. 2.   Feature Importance using XGBoost (before the exam).
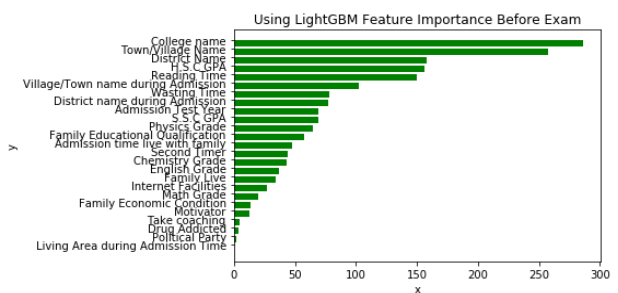

Fig. 3.   Feature Importance using LightGBM (before the exam).
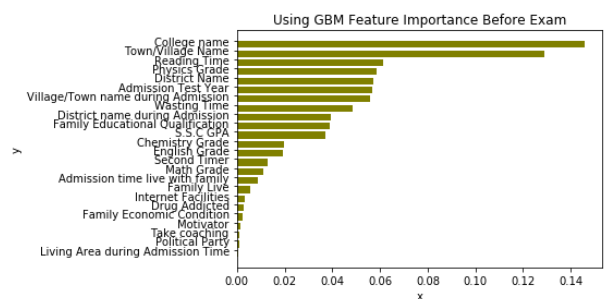

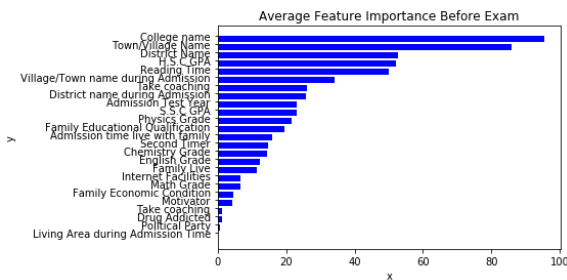Fig. 4.   Feature Importance using GBM (before the exam).

Fig. 5.   Average Feature Importance (before the exam).

Here, the academic performance of a candidate e.g. H.S.C GPA and Reading time is an obvious reason for better chances of admission. As we mentioned in Section III, since Bangladesh is a developing country and the development of the country is not equally distributed based on location the address and nativity of the candidate i.e. College name Town/Village name, District name are discovered as important features to predict undergraduate admission before the examination.

D. Feature Importance after the Exam

Fig. 6 shows the important features for predicting the admission opportunity of the applicants after the test using the XGBoost learning algorithm. Fig. 7 shows the important features for predicting the admission opportunity of the applicants after the test using the LightGBM learning algorithm. Fig. 8 shows the important features for predicting the admission opportunity of the applicants after the test using the GBM learning algorithm.

The authors combined the feature importance of the aforementioned three learning algorithms. This study did the average feature importance (score) of each feature and plotted it in Fig. 9. It shows the features that affect most for getting an opportunity of admission in engineering faculty after participating in the admission test and these are as follows:

- Obtained marks

- Admission test year

- Town/Village name

- College name

- H.S.C GPA

Here, the academic performance of a candidate e.g. Obtained marks in the admission test and H.S.C GPA is understandable as the important features to predict undergraduate admission after the examination. Along with these, the address and nativity of the candidate i.e. Town/Village name and College name are discovered as important features because the development of Bangladesh is not equally distributed based on geographical factors. The students who live in town usually have more facilities than rural students. Admission test year is also found as an important feature because it is observed that some years' questions were comparatively easier and some years' questions were comparatively hard which may affect getting a chance in the admission.
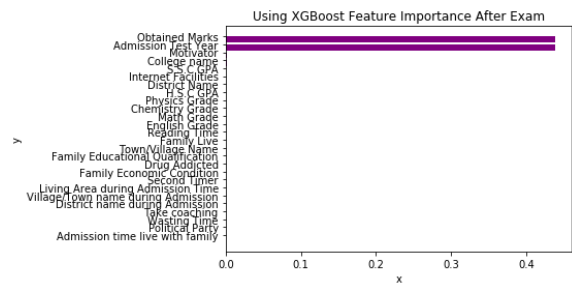

Fig. 6.   Feature Importance using XGBoost (after the exam).
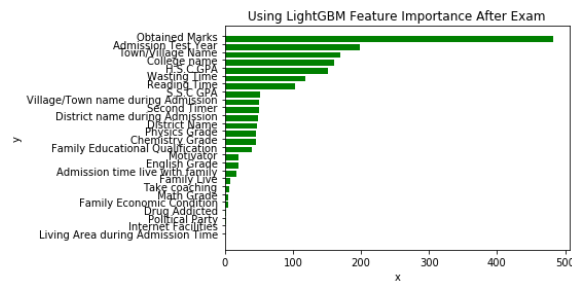

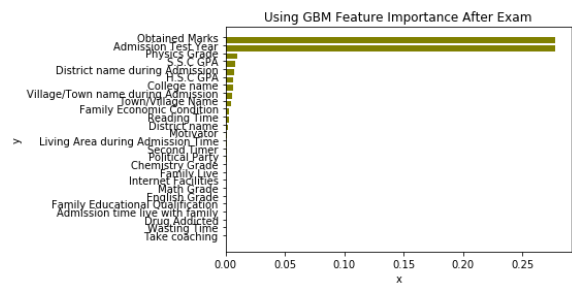Fig. 7.   Feature Importance using LightGBM (after the exam).

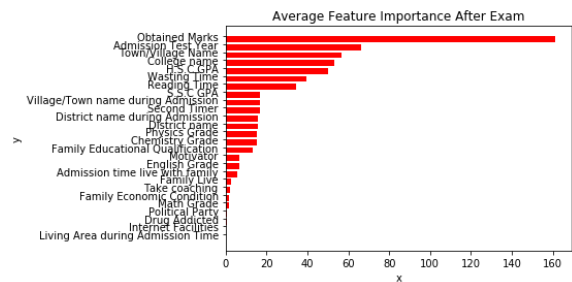Fig. 8.   Feature Importance using GBM (after the exam).

Fig. 9.   Average Feature Importance (after the exam).

V.   DISCUSSION

The accuracy, F1 score results are different while using different learning algorithms as showed in Table II and Table III. Fig. 10 also shows a comparison among these three algorithms using a bar chart plotting for predicting admission before the test. XGBoost gives the best accuracy and F1 score among the three algorithms although the score is lower as expected. In this case, the dataset did not include the expected score of the applicants in the examination. Fig. 11 shows algorithm comparison for predicting admission after the test. Unlike the first case, the score is achieved higher i.e. above 80%. Here, GBM outperforms XGBoost and LightGBM. The

evaluation metrics – accuracy and F1 score both are 95% for GBM. The proposed model using LightGBM and XGBoost also shows excellent results i.e. 93% and above 80% respectively. Hence, the proposed model predicting undergraduate admission after the examination must be very efficient and effective for the students.
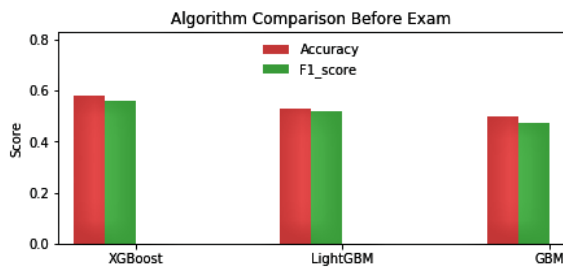

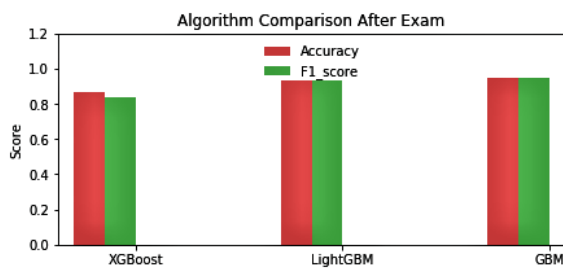Fig. 10. Algorithm Comparison before the Admission Test.

Fig. 11. Algorithm Comparison after the Admission Test.

VI. CONCLUSION

In this research, the authors used three boosting techniques to estimate the probability of getting an undergraduate admission in the engineering faculty at BSMRSTU, Bangladesh. The root of the dataset is the students of BSMRSTU who are currently studying in the engineering departments and who are not in the engineering departments. Using some machine learning techniques, the authors developed two models separately – the admission predictive model before the admission test and the admission predictive model after the admission test. The authors extensively investigate and analyzed these models. The evaluation results show that the proposed model can able to assist the students in predicting admission opportunities. This study performed the prediction only for the engineering unit at BSMRSTU. This method can be applied to predict admission in any other faculties or universities also.

REFERENCES

[1] "The Daily Star," 2017. [Online]. Available: https://www.thedailystar.net/backpage/public-universities-admission-still-uphill-battle-1438285. [Accessed 2020].

[2] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, 3 ed., San Francisco:Morgan Kaufmann Publishers, 2011.

[3] Aggarwal and C. C, Data mining: the textbook, Springer, 2015.

[4] C. C. Aggarwal, Data Mining: The Textbook, 1 ed., Springer International Publishing Switzerland, 2015.

[5] T. Calders and M. Pechenizkiy, "Introduction to the special section on educational data mining," SIGKDD Explorations, vol. 13, no. 2, pp. 3-6, 2012.

[6] J. P. Bigus, Data mining with neural networks: Solving business problems from application development to decision support, McGraw-Hill Companies, 1996.

[7] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601-618, 2010.

[8] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," ArXiv Preprint ArXiv:1201.3417, 2012.

[9] El-Halees and A. M, "Mining students data to analyze e-Learning behavior: A Case Study," Mining students data to analyze e-Learning behavior: A Case Study, vol. 29, 2009.

[10] C. Romero, S. Ventura and E. Garcia, "Data mining in course management systems: Moodle case study and tutorial," Computers & Education, vol. 51, no. 1, pp. 368-384, 2008.

[11] P. Binu, A. Chandran and M. Rahul, "A Cloud-Based Data Analysis and Prediction System for University Admission," in 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 2019.

[12] M. S. Acharya, A. Armaan and A. S. Antony, "A comparison of regression models for prediction of graduate admissions," in International Conference on Computational Intelligence in Data Science (ICCIDS), 2019.

[13] A. M. Roa, N. Dharani, A. S. Raghava, J. Buvanambigai and K. Sathish, "College Admission Predictor," Journal of Network Communications and Emerging Technologies (JNCET), vol. 8, no. 4, April 2018.

[14] B. Ghai, "Analysis & Prediction of American Graduate Admissions Process," 2018.

[15] N. Gupta, A. Sawhney and D. Roth, "Will I get in? modeling the graduate admission process for American universities," in IEEE 16th International Conference on Data Mining Workshops (ICDMW), 2016.

[16] R. V. Mane and V. Ghorpade, "Predicting student admission decisions by association rule mining with pattern growth approach," in International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2016.

[17] A. B. Raut and A. A. Nichat, "Students Performance Prediction Using Decision Tree Technique," International Journal of Computational Intelligence Research, vol. 13, no. 7, pp. 1735-1741, 2017.

[18] P. M. Arsad and N. Buniyamin, "A neural network students' performance prediction model (NNSPPM)," in International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), IEEE, 2013.

[19] Ş. Z. Erdoğan and M. Timor, "A data mining application in a student database," Journal of aeronautics and space technologies, vol. 2, no. 2, pp. 53-57, 2005.

[20] A. Ktona, D. Xhaja and I. Ninka, "Extracting Relationships between Students' Academic Performance and Their Area of Interest Using Data Mining Techniques," in Sixth International Conference on Computational Intelligence, Communication Systems and Networks, 2014.

[21] T. Devasia, V. T. P and V. Hegde, "Prediction of Students Performance using Educational Data Mining," in International Conference on Data Mining and Advanced Computing, 2016.

[22] J. Ruby and K. David, "Analysis of Influencing Factors in Predicting Students Performance Using MLP-A Comparative Study," International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, no. 2, pp. 1085-1092, 2015.

[23] A. A. Aziz, N. H. Ismail, F. Ahmad and H. Hassan, "A Framework for Students' Academic Performance Analysis using Naïve Bayes Classifier," Jurnal Teknologi, vol. 75, no. 3, 2015.

[24] C. Anuradha and T. Velmurugan, "A Comparative Analysis on the Evaluationof Classification Algorithms in the Prediction of Students Performance," Indian Journal of Science and Technology, vol. 8, no. 15, 2015.

[25] P. Kaur, M. Singh and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," in International Conference on Recent Trends in Computing, 2015.