Abstract: For a student higher study is a new chapter of his study. For a majority of us, studying at a reputable university abroad is an imminent goal. For a long time, students from developing countries tend to seek higher study from developed countries. Higher educated peoples are played an important role in building a civilized society and a developed nation. The more educated a nation is, the more developed it is. Almost all the countries are trying to increase the quality of their higher education day by day. Although Bangladesh is a developing country, Bangladesh is also becoming the preferred country for foreign students' day by day. The purpose of this paper is to Analysis foreign students' academic and lifestyle data to identify whether Bangladesh is suitable for higher education or not, those who are currently studying here are satisfied or dissatisfied. If they have any problems then what kind of problems are they experiencing and how to resolve them. we have identified the university's problems where they need to improve to attract more foreign students. For the case of foreign students, we have collected data from 399 present international students in Bangladesh. we have used some popular classification algorithms such as support vector machine, KNN, ANN, Random Forest. We also present a comparison of used machine learning algorithms over different evaluation metrics. After all, we have analyzed the data to show which factors play a major role in students' decision making

Introduction: The educational process encourages the full development of a person's inherent qualities and helps him to acquire the skills needed to establish himself as a productive member of society. Education is the acquisition of knowledge or skills in the general sense. In a broad sense, education is the process of gaining systematic knowledge and education is a continuous practice of developing to the fullest of potential. According to Article 17 of the Constitution, every child in Bangladesh has the right to free education up to the secondary level. Secondary and higher secondary level educational institutions are affiliated to ten (10) education boards. The boards conduct three public examinations: Junior School Certificate (JSC) examination, Secondary School Certificate (SSC) examination and Higher Secondary Certificate (HSC) examination. Educational institutions at the higher secondary level are known as colleges. From then on, the journey of higher education of Bangladeshi students is started.

Higher education usually involves a degree-level or degree qualification. In most developed countries, the population (up to 50%) now enters higher education at some point in their lives [Wikipedia]. Higher education is very important for the formation of a civilized nation and the national economy. Civilized society has been relying on higher education institutions for many centuries to create leaders in various professions. The scope of higher education is expanding over time and will continue. There has been remarkable progress in various sectors (business, commerce, industry, and communication) centered on the higher educated people. That's why higher education institutions must be aware of the quality of education provided in these areas. Therefore, higher education in the modern world must be developed in many special disciplines and it should be suitable for a large number of students.

After World War II, there is a tendency for students from poor and developing countries to seek their higher education from developed countries [18]. Now the question is why? Every student

wants to go from his current education system to a good education system. The reason is there is a big difference between the education system of the developed country and the education system of the developing country. There are many opportunities for research abroad. Those who study abroad are far ahead of the general public in terms of independence, self-reliance, intelligence and creative ability. As leaving their family and living alone in a foreign country, they are self-reliant and skilled in solving problems. Not only that, they are not lacking in any challenging task. By staying abroad, a student will become aware of the politics, culture, customs of that country. In addition, he will have a different view of the world. Learn about the world's political issues Being abroad, he has to learn the language of that country, which will add another pastor to the crown of his experience. Higher education abroad will not only improve the educational qualifications and values of a student, but also increase his professional skills.

Before 1991 there had only eight public universities in Bangladesh. In spite of having the knowledge, there had no opportunity to seek higher study for the limited number of sits. Therefore, the government of Bangladesh took the initiative to set up public universities as well as private universities so that Bangladeshi students can get higher education in the country. After 1991, the private university of Bangladesh started its journey. At present, 42 public universities and 109 private universities in Bangladesh have confirmed the quality of their higher education (UGC report 2018). Currently, these universities are studying with Bangladeshi students as well as foreign students (especially South Asia, Africa). In Bangladesh, the number of foreign students is increasing day by day. But after talking to the current foreign students, we found out that due to some problems (could not attend the desired university, cannot afford the cost) students go back to their country without completing their studies. This threatens the student’s carrier. Because the student leaves the country for higher education and spends a lot of money behind it. Therefore, one of the challenges of the present government of Bangladesh is to ensure the safety of foreign students and how to ensure the quality of higher education. The structure of the paper is organized as follows.

In section 2 there has been discussed related work. In section 3 there has been shown the data distribution. In section 4 there has been shown the Methodology. In section 5 there has been discussed the result and analysis. In section 6 there has been shown conclusion and references of the paper.

Related Work: The process of internationalization in higher education is progressing steadily [16].

GEORGE V. COELHO at all [1]. In their research, they described the main factors that put foreign students at risk in higher education as a result of the transfer of new cultures and environments. They Mentioned Many complications arise when young people migrate abroad for higher education as they become acquainted with new cultures and environments. So, they have to adapt to that environment as soon as possible to overcome this complication. They highlighted a life cycle of the students and they said that child's first school experience, the junior school transition, the transition from high school to college, graduate student stress, etc. Education overseas may be a major developmental and psychosocial move in a foreign student's life. Like other moves, it

represents an arrangement of stages of high-risk situations that deliver emotional push as well as openings for adapting behavior.

C. D. THROSBY at all [7]. They have shown in their research how foreign students play an important role in the economic development of the host country. They described both direct and indirect costs-benefit. They have mentioned tuition fee, accommodation, meals and the costs of any additional services provided, such as special counselling, or host country language courses as a direct benefit. for the indirect benefit, they point out that local students can become acquainted with a wide variety of cultures, and that foreign students take the research sector further by participating in research. Similar to these benefits There are some possible parallel costs such as enmity from local students, unpleasant reaction by foreign students to the host institution, etc.).

There have been a few endeavors to measure the total benefits and costs of foreign students in host countries (for example, Reubens 1975[3]; Blaug 1981[4]; Jenkins 1983[5]; Winkler 1983[6]; Chishti 1984[7]; Fry 1984[8]; Manning et al.

Every country follows some foreign policy for foreign student. According to [9] Prior to 1986, about 45 to 60 per cent of foreign students in Australia were studying on scholarships. As a result, the Australian government had to pay a large subsidy every year. In 1986, the government of Australia issued a proclamation to all educational institutions in their country, asking all educational institutions to declare their course as fully paid for foreign students. As a result of the Australian Government's decision, students from many developing countries, despite their qualifications, were barred from pursuing higher education in Australia. To address this problem, the Australian government introduced the Merit Scholarship in 1990. Within 4 years of this rule, the number of foreign students in Australia increased from 24,000 to 55,000, of which 30,000 were fully paid.

Nowadays machine learning algorithms have become very popular in research. At present machine learning techniques are being used to solve other problems as well as the problems of educational institutions such as institutions, teacher quality (Góes et al. 2014[10]" Education quality measured by the classification of school performance using quality labels". Xiao-YanLiu 2015[11]" Private colleges teachers evaluation system based on support vector machine (SVM)."), examination and assessments (Muklason et al. 2017[12]") Fairness in examination timetabling: student preferences and extended formulations."), measure practices impact (Delen et al. 2013[13]" A comparative analysis of machine learning systems for measuring the impact of knowledge management practices"), learning product selection (Alptekin and Ertugrul 2010[14]" An integrated decision framework for evaluating and selecting e-learning products"; Oztekin et al. 2013[15]" A Machine learning-based usability evaluation method for eLearning systems."), course planning (Abdahllah 2015[16]" A decision support model for long-term course planning") and more.

According to Acharya A, Sinha D (2014) [17]at all. Developed a model using machine learning to predict student's performance. They have mentioned the academic performance of students depends on previous academic records, economic status, family background, performance in mid semester examination. Based on these factors they apply Decision Tree (DT). Bayesian Networks (BN), Artificial Neural Networks (ANN), Support Vector Machines (SVM).

Tan M, Shao P (2015) by using Artificial Neural Network (ANN), Decision Tree (DT) and Bayesian Networks (BNs) developed a model to predict the dropout of student in eLearning program.

Although some of these studies addressed educational problem issues at the higher education level, none have considered the satisfaction and dissatisfaction on host country. While previous research focuses on prediction, the most commonly used algorithms are artificial neural networks and support vector machines. Some authors, including ourselves, go further into comparing these algorithms' performance.

3 Methodology:

3.1 Proposed method

- Collecting real field data.
- Data pre- processing.
- Train Model.
- Test Model with test data.
- Apply Machine Learning Techniques: -
  - SVM
  - ANN
  - KNN
  - Decision Trees
  - Random Forest
- Find predicting accuracy.

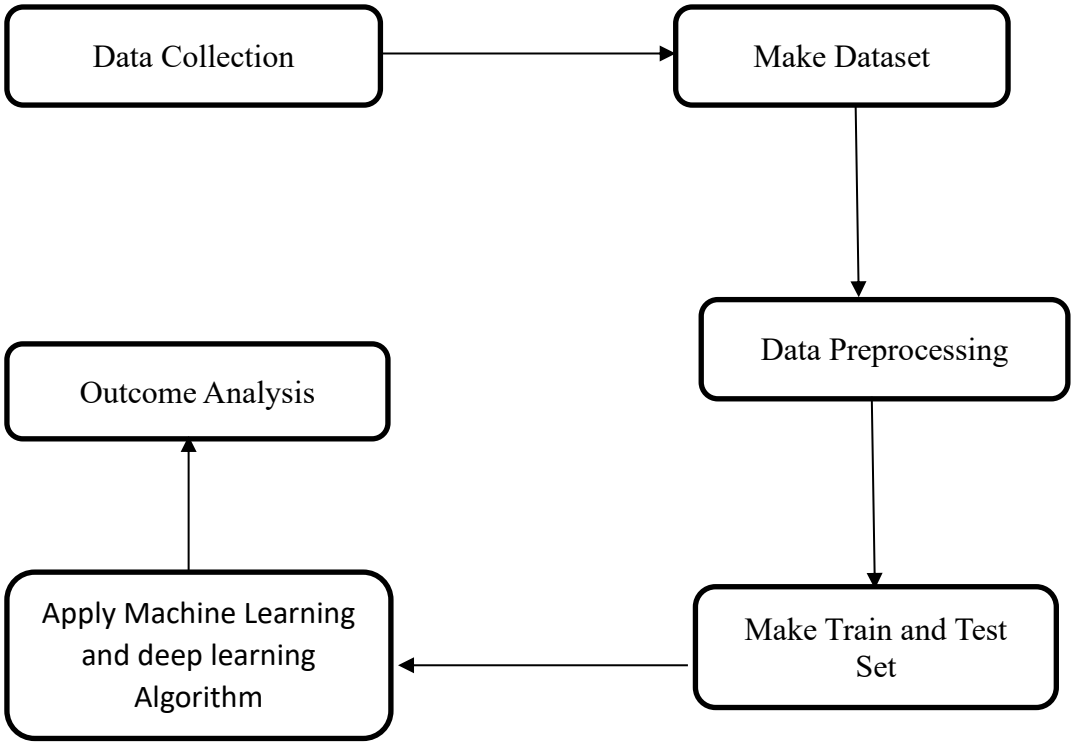

Figure 1: Proposed method

3.2 Data collection

Data have been collected from foreign students studying in Bangladesh universities, such as (Daffodil University, Independent University, Dhaka International University, City University, Asian University, IUBAT University, and so on). The dataset contains the current situation of foreign students. This dataset contains 25 columns where each column speaks to a special piece of data about a foreign student. Since every column provides valuable information for satisfaction or dissatisfaction of foreign students. 25 columns have been selected in this paper, these columns contain both "satisfied" and "dissatisfied" Foreign students' information. The dataset contains 399 rows and 25 columns based on the following questions which are shown below.

1. It's easy to visa processing from your country to Bangladesh?

2. After coming to Bangladesh are you face any complexity of visa?
3. Is it easy to send money from your country to Bangladesh?
4. Do you feel safe in Bangladesh?
5. Has the government of Bangladesh given you any insurance?
6. Do you satisfy with the accommodation in Bangladesh?
7. Bangladeshi classmates are friendly?
8. Are Bangladeshi teachers friendly and helpful?
9. Are General people of Bangladesh friendly?
10. Are Bangladeshi police helpful?
11. Do you have to pay any tax in Bangladesh?
12. Can you use ATM card in Bangladesh?
13. In Bangladesh are you a victim of racial decimation?
14. Do you face any problem with food?
15. Do you face any problem with changing climate?
16. Do you face any political problem?
17. Do you face any corruption?
18. Do you face any problem with the Bangladeshi education curriculum?
19. Are you satisfied with the health service of Bangladeshi hospitals?
20. Do you enjoy the festivals of Bangladesh?
21. Do you face any problem on the first meeting?
22. Can you buy any vehicle in Bangladesh easily?
23. Is there any problem to celebrating the own religion?
24. Do you face sexual harassment?
25. What is your gender?

3.3 Ground Truth Dataset

This segment has been talked about the method utilized to develop ground truth label information (on whether the foreign students are satisfied or dissatisfied). the dataset has been split into two sets: - (1) for the positive side is taken '1' means satisfied student and (2) for the negative side is taken '0' means dissatisfied student. 399 records have been analyzed. Among the whole dataset,

80% has been gotten 1 and 20% has been gotten 0, so it means the majority of the respondents were satisfied.

Total number of students	399
Satisfied students	279
Dissatisfied students	120

Table 1: Data set

3.4 classification model

The machine learning classification has two steps, one is the learning step and the other is the prediction step. Learning data is provided to teach the machine and data is provided to predict the prediction step. This stage develops a prediction model for satisfaction/dissatisfaction recognition, by considering the features as input. Considering our preparing corpus B = p_1, p_2.....p_n on n students' data, such that each information p_i is labeled with the class either satisfaction or dissatisfaction, where L = $l_1|l_2$. The task of classifier *f* is to discover the corresponding label for each student data.

$$f: B \in L \qquad f(p) = l$$

Four well-known classifiers have been used in this paper: support vector machine (SVM), Decision Tree (DT), Random Forest (RF), k-nearest neighbor (KNN).

3.4.1 Artificial Neural Network

Network means basically a set of nodes where node is connected to each other in some way (directly or through anything). A neural network is a network of neurons where neurons are interconnected and can exchange information with each other. The neurons are arranged in one or more layers. The calculation of data is done according to the layer and information is exchanged from one layer to another. Below is a picture of a very common neural network.

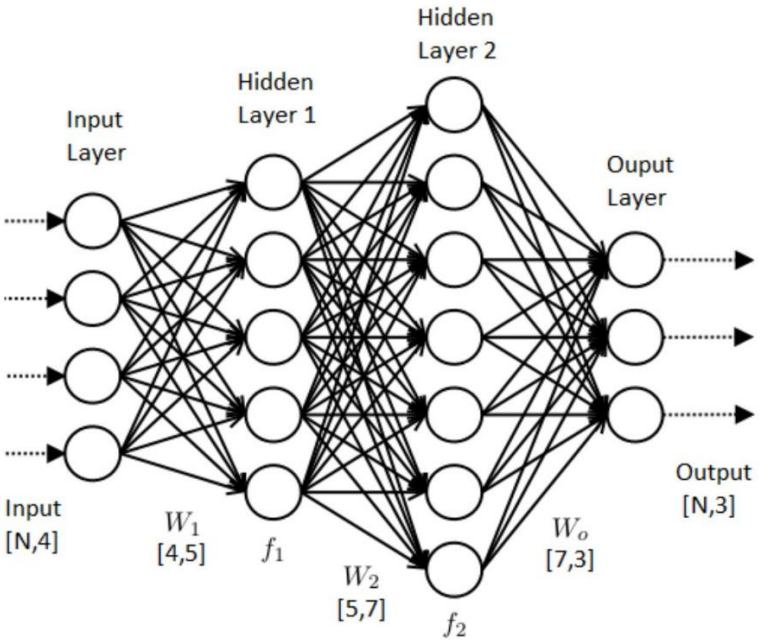

Figure 2: Artificial Neural Network

The layers are classified into three categories namely the input layer, hidden layer, and output layer. The connections between neurons connection are assigned specific weight. The connections between neurons correspond to a few weights. So, training a neural network involves altering all those weights such as in the case given an input after performing all the calculations it gives the proper output. In this backpropagation algorithm first input is gone in one way to output and after that propagate back from the output to input adjusting weights. There's too a bias unit whose weights are balanced. And at last, an activation function is used to get the proper output according to the training set. The backpropagation is rehashed until the error is decreased to a really little esteem. The multilayer perceptron is more advantageous with features like non-linear mapping and noise tolerance. It is more used in data mining since of its great behavior with respect to predictive knowledge

3.4.2 Support vector machine:

Support vector machine is supervised learning model. It is used in classification and regression machine learning problems. The main goal of SVM is to separate the classes by drawing a hyper line with the highest margin between the two classes. Notice the figure below, the points of the two classes closest (‘solid round’ ‘solid square) of them are called support vectors. The distance

between these support vectors is called the margin. The higher the margin, the less likely the points are to be misclassified. Because the closer the points of two different classes are, the more likely they are to overlap. Supported vector help to decipher boundary and margin. Other points of a class are not as important as the classification, except for the support vectors. If there are two hyper plans in which A can properly classify the two-class, but the margin is very low. On the other hand, the margin of B is too high but there are some errors. A will be selected. This higher dimensional dividing is known as the SVM kernel and can be defined by any mathematical surface. A few of the more common kernels are linear, quadratic, polynomial and Gaussian radial basis function.

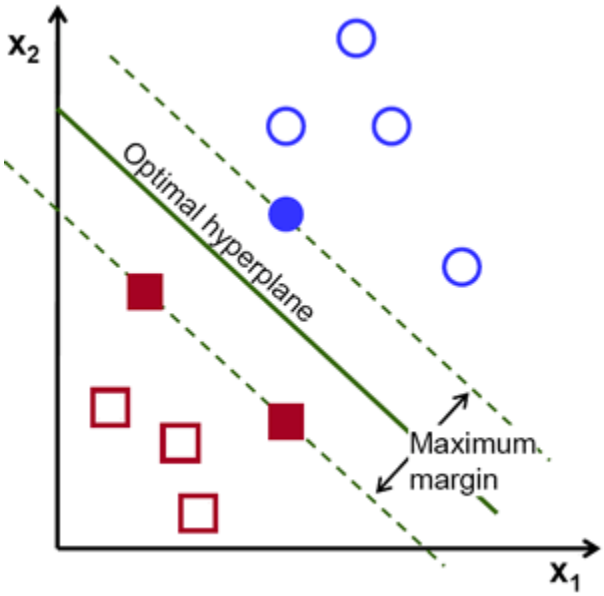

Figure 3: Support Vector Machine

3.4.3 K-Nearest Neighbors

K-Nearest neighbor algorithm (k-NN) is a non-parametric and lazy (training is not required) method. It is a supervised machine learning algorithm. K-Nearest neighbor algorithms (k-NN) can solve both classification and regression problems. The ‘k’ in KNN algorithm is the number of nearest neighbors taken into consideration. In this model, the value of k has determined by the square root of the total data.

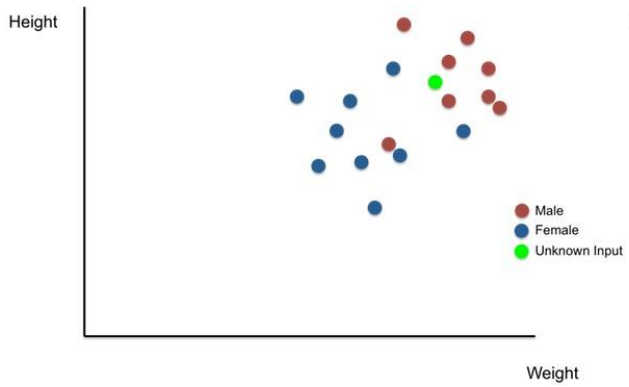

Figure 4: Before k-NN Classification

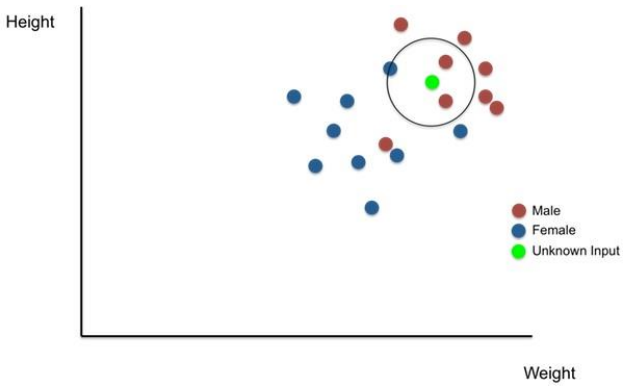

Figure 5: After k-NN Classification

Figures 4 & 5 show the scenario before k-NN classification and after k-NN Classification.

Here, 'unknown' has to be put in one of the two classes – 'Male or 'Female. With Unknown input as center and radius sufficient to encircle 3 nearest neighbors (k=3), a circle is drawn. Class of unknown input is defined based on the majority.

The four main step of the algorithm are: -

- All other data points must be calculated from unknown data points.
- Ascending or descending orders should be resorted according to the distance standard.
- The first k is to take the data points from the sorted data points.
- Unknown data point should be identified in that class as the number of data points in which this class has the highest number of times.

3.4.4 Decision Tree

Decision tree is supervised learning algorithms. It can be used for solving regression and classification problems too. In this algorithm, data is split by a particular node. a child note is created after each value is split. If the subset of each child node is the target value, then it will stop otherwise recursive split will occur. This algorithm uses an algorithm called ID3 to select the root note and adjust the order of intermediate notes.

The id3 algorithm uses the greedy method approach to get a decision tree that returns the highest information gain or minimum entropy.

Entropy: entropy is, if a dataset is partitioned against a feature, then how well that partitioning can partition the target variable's column

$$E(S) = \sum_{i=1}^{c} -P_i \log 2\ P_i$$

Gain: Information gain is based on the decrease in entropy after a dataset is

Split on an attribute.

$$Information\ Gain = Entropy\ (before) - \sum_{j=1}^{k} Entropy(j,\ after)$$

3.4.5 Random Forest:

Decision tree is supervised learning algorithms. It can be used for solving regression and classification problems too, it is mainly used for classification problems. Forest means the sum of many trees. Similarly, random forest is the sum of many decision trees.

Random Forest completes its tasks by following the steps below: -

- To begin with the selection of random samples from a given dataset.
- Following, this algorithm will develop a decision tree for each sample. At that point, it will get the prediction result from each decision tree.
- At the last stage, it will take result depending upon the majority voting.

performance evaluation

In this segment, some performance metrics have been used to evaluate the performance of classifiers such as Recall, False Positive rate (FPR), F-measure, Accuracy, and Precision. A confusion matrix is an easy way to calculate algorithm performance. Where the output can be of two or more types of classes. confusion matrices are two-dimensional matrices, dimensions are 'Actual', 'Predict' and more, each dimension have “True Positives (TP)”, “True Negatives (TN)”, “False Positives (FP)”, “False Negatives (FN)” as shown below –

Actual Class	Predicted Class	
	Negative	Positive
Negative	TN	FP
Positive	FN	TP

Accuracy means calculating how many accurate predictions an algorithm can make (TP and TN).

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

Precision means the ratio of the number of positive predictions to the number of positive predictions the algorithm ha made.

$$Precision = \frac{TP}{TP + FP}$$

Recall is the ratio of true positives to the cases that are actually positive. It is the percentage of corrected cases that are selected.

$$Recall = \frac{TP}{TP + FN}$$

F-measure is the harmonic mean of precision and recall.

$$F = 2\frac{Precision * Recall}{Precision + Recall}$$

5 Results and Discussion:

The model has been built on current foreign students in Bangladesh. The dataset has been split into 60/40, in which 60% of the data set has been used for the training model and 40% of the data has been used for the testing model. five popular classification algorithms have been used, such as SVM, ANN, KNN, DT, RF. In KNN classification, the value of k has been used 5. The training model is built with 60% of the dataset. and KNN produces 80% Accuracy. 60% of data has been used to train SVM. The kernel has been used linear because the data set has only two classes one is "satisfaction" other is "dissatisfaction". SVM produces 95% accuracy. Next, the Artificial neural network model is built using a multilayer perceptron and it also used 60/40 split of data. 4 hidden layers have been used. Among the five algorithms, ANN has given the highest accuracy it produces

98% accuracy. the remaining two algorithms decision tree and random forest give 92% and 94% accuracy, respectively.

Table 2: Performance of Classifiers

Method	Precision	Recall	F_1 value	Accuracy
SVM				99%
KNN				89%
DT				92%
RF				94%
ANN				97%

Table 2: Performance of Classifiers

6 Conclusion:

In this study we have collected data from current foreign students and analyze what kind of problems they are facing and how to solve these problems. are they satisfied or dissatisfied with the Bangladesh educational curriculum? we have collected data from foreign students studying is Bangladesh. five popular classification algorithms have been used on the algorithm. According to the results support vector machine classifier works better among all the other classifiers for this dataset. in this paper, we have collected data from the students of only two countries (Somalia, Uganda) studying Bangladesh. in the future, we will collect a large amount of data form all countries students who studying in Bangladesh. and also, we will try to apply more classification algorithms for better results.

Reference:

1. GEORGE V. COELHO, Ph.D. “THE FOREIGN STUDENT'S SOJOURN AS A HIGH-RISK SITUATION: THE "CULTURE-SHOCK" PHENOMENON RE-EXAMINED” Uprooting and Surviving, 101-107.
2. C. D. THROSBY.” The financial impact of foreign student enrolments” Kluwer Academic Publishers research on Higher Education 21: 351-358, 1991.
3. Reubens, E. P. (1975). 'The new brain drain from developing countries: international costs and benefits, 1960-1972', in Leiter, R. D. (ed.), Costs and Benefits of Education. Boston: Twayne Publishers, pp. 178-215.
4. Blaug, Mark (1981). 'The economic costs and benefits of overseas students', in Williams, Peter (ed.), The Overseas Student Question. Studies for a Policy. London: Heinemann, pp. 47-90.
5. Jenkins, H.M. (1983). 'Economics: analysing costs and benefits', in Jenkins H.M. et al. (eds.), Education of Students from Other Nations. San Francisco: Jossey Bass, pp. 237-250.

6. Winkler, D. R. (1983). 'The costs and benefits of foreign students in US higher education', Journal of Public Policy 4 (2), 115-138.

7. Chishti, S. (1984). 'Economic costs and benefits of educating foreign students in the United States', Research in Higher Education 21 (4), 397--414.

8. Fry, Gerald (1984). 'The economic and political impact of study abroad', Comparative Education Review 28, 203-220.

9. THE AUSTRALIAN DEPARTMENT OF EMPLOYMENT, EDUCATION AND TRAINING." Programmes and policies for foreign students in Australia" country report on Australia Higher Education 21: 379-388, 1991.

10. Góes ART, Arns Steiner MT, Steiner Neto PJ (2014) Education quality measured by the classification of school performance using quality labels. Appl Mech Mater 670:1675–1683.

11. Xiao-YanLiu (2015) Private colleges teachers evaluation system based on support vector machine (SVM). In: International Conference on Applied Science and Engineerin Innovation ASEI 2015, no. Asei, pp. 1918–1921.

12. Muklason A, Parkes A, Ozcan E (2017) Fairness in examination timetabling: student preferences and extended formulations. Appl Soft Comput 55:302–318.

13. Delen D, Zaim H, Kusey C (2013) A comparative analysis of machine learning systems for measuring the impact of knowledge management practices. Decis Support Syst 54:1150–1160.

14. Alptekin E, Ertugrul K (2010) An integrated decision framework for evaluating and selecting e-learning products. Appl Soft Comput 11:2990–2998.

15. Oztekin A, Delen D, Turkylmaz A (2013) A Machine learning-based usability evaluation method for eLearning systems. Decis Support Syst 56:66–73.

16. Abdahllah M (2015) A decision support model for long-term course planning. Decis Support Syst 74:33–45.

17. Anal Acharya, Devadatta Sinha 2014 "Early Prediction of Students Performance using Machine Learning Techniques" International Journal of Computer Applications (0975 – 8887) Volume 107.