Prediction of Student's Performance Using Random Forest Classifier

Sourav Kumar Ghosh and Farhatul Janan
Department of Industrial and Production Engineering
Bangladesh University of Textiles
Dhaka 1208, Bangladesh
sourav@butex.edu.bd & janan_ipe@butex.edu.bd

Abstract

Measuring student performance based on both qualitative and quantitative factors is essential because many undergraduate students could not be able to complete their degree in recent pasts. At present, students' dropout rate in university is gradually increasing and many bright students sometimes just cannot cope with the universities. The first-year result of a student is very important because in the majority of cases this drives the students to be either motivated or demotivated. So, the first-year student performance of a renowned university in Bangladesh is investigated in this paper. This research is mainly based on finding the reasons for students' different types of results and then predicting students' performance based on those eleven significant factors. For this purpose, a popular supervised machine learning algorithm, random forests (RF) have been used for classifying students' different levels of results and predicting students' performances. The input dataset for both training and testing were taken by merging the values obtained from two surveys done on students and experts using fuzzy ANFIS analysis. The result exhibits that RF can perform the classification of multiple classes based on many distinguishing features with 96.88 percent accuracy. This proposed model can also be applied to predict course-wise students' performances and its precision can also be greatly improved by adding new factors.

Keywords

Random Forest Classifier, Fuzzy ANFIS, Performance Prediction, Classification

1. Introduction

Education is one of the basic needs of human beings. It has many levels such as early childhood education (level 0), primary education (level 1), lower secondary education (level 2), upper secondary education (level 3), postsecondary non-tertiary education (level 4), short-cycle tertiary education (level 5), bachelor's or equivalent level (level 6), master's or equivalent level (level 7) and so on. Dropout from any educational stage is also a common phenomenon. However the dropout rate increases as moving to the upper stages especially at the undergraduate level. There are many factors which may play a vital role in increasing the dropout rates. In the majority of the cases, when students' results start degrading that is the main trigger point for increasing the probabilities of those students to be dropped out. This research is basically focused on finding different factors that affect freshmen students' results at the undergraduate level.

University is a very important place for students where they shape their futures. It creates opportunities for students to fulfill their dreams and to achieve a purpose in life. Sometimes, the reverse case also occurs. After coming to university, many students cannot adapt themselves to the university's environment for study where some other causes like some students become more involved in extra-curricular activities or students' politics, some start to avoid studies because they dislike their departments, some do not like the career they have ahead, etc. Due to these various known and unknown reasons, students' performances in the university in many cases tend to be low which in turn affects their results. The trend shows that most of the poor performances are yielded from the early-stage failure especially in the 1st year of the study. So, the performance of 1st-year students should be analyzed to find the real root cause of the problem related to student's performance.

So, our main motivation behind this work was to help students understand the attributes which are responsible for their poor results so that they can improve their performance. If the major factors are identified and monitored, it will give the students, course teachers, and the administrations to ameliorate the study environment. On the other hand, if students can anticipate the reasons for degrading their results thus, they can work on those to improve their performances. Additionally, the institution (in this case university) itself can take some proper steps to improve students' performances according to this study. We have performed our experiment at Bangladesh University of textiles which is the best public university in the textile sector in Bangladesh. The freshmen students are the top students in their previous educational level and have to go through an entrance exam to get the chance to study in this university. The results of the recent first year (45th batch) students of this university were taken for this research. For identifying the factors, we have done surveying on students and on some experts (here experts mean different course teachers associated with students). After completing the survey, factors and their ratings were identified which showed the reasons for students' different levels of performance. We have merged the ratings of students and experts on a particular factor on a particular student using fuzzy ANFIS analysis for using these modified data as input to the model. 80 percent of these data were used for training and the remaining 20 percent was used for testing (predicting students' results based on those factors) and finally by computing the accuracy of the model, the validity of the formulation of the model (factors' identification and their ratings) was accomplished. In this model raw data are modified using expert opinions and fuzzy logic. For this reason, the model performed better than other traditional models which only considered raw data.

1.1 Objectives

In this paper, we have formulated the problem as a multi-class classification problem. Different techniques are available to solve this problem. After analyzing the research works on classification problems and prediction of students' performances, it was found that Random Forest (RF) classifier was used in many works and its performance was better than other methods. So, our aim was to implement this method to predict students' performance by introducing fuzzy logic which had helped to merge different sets of input databases under certain rules to a single input database. The random forest consists of many decision trees. Each individual decision tree gives one class prediction and the class with the highest number of votes becomes the model's final prediction.

2. Literature Review (12 font)

Over the years, many researchers have already done many works regarding analyzing the reasons behind a student's performance. Some of these research works include only analyzing the attributes which have direct or indirect effects on students' performance and some also include predicting other students' results based on the studied attributes using different learning algorithms. Most of these learning algorithms are artificial intelligence (AI) techniques such as heuristics, machine learning, artificial neural networks, random forest, Bayesian classifiers, etc.

Naive Bayesian (NB) data mining technique was applied to predict the student performance based on 19 attributes such as gender, food habit, the medium of teaching, family status, family income, students' grade, and son on [1]. Classifications algorithm such as the decision tree algorithm (J48), NB, Bayesian classifiers, k Nearest Neighbor (KNN) algorithm, and two rule learner's algorithms (OneR and JRip) were used to estimate the student's performance. The overall accuracy of the NB classifier was the greatest among other classifiers [2]. Factor reduction was implemented by correlation-based feature selection (CBFS), chi-square-based feature evaluation (CBFS), and information gain attribute evaluation (IGATE). The decision tree (DT) algorithm worked more efficiently than other machine learning (ML) algorithms on a case study of some undergraduate students in Kolkata [3]. First-year bachelor student's data on a particular course over 8 years was analyzed with the help of DT, NB, and rule-based classifiers. It was posited (found) that a rule-based classifier gave the best prediction results compare to the other two classifiers [4]. Frequent pattern tree algorithm, ensemble semi-supervised learning (SSL) algorithm, recurrent neural network (RNN) and DT techniques were employed to forecast student's academic performance [5][6][7]. Socio-economic factors and entrance examination results were used as the primary factors to predict the student's cumulative grade point average (CGPA) by applying ANN with the Levenberg–Marquardt algorithm [8]. The two-level classification algorithm was implemented to calculate the expected graduation time where the passed or failed students were differentiated in the first level and three different periods of graduations were classified in the next level [9]. A literature review of various techniques applied to predict student performance was studied. Techniques were grouped into three clusters such as fuzzy logic, data mining techniques, and hybrid [10]. Sequential Minimal Optimization (SMO) and NB were combined into a hybrid model, SM Naive Bayes (SMNB) model was developed to predict the student performance. This hybrid model was surpassed by other individual models in accuracy [11]. Student performance was forecasted by random forest algorithm considering e-learning environment where lab total, assignment submission, mid-term was assessed as the principal attributes [12]. RF classifier outperformed C4.5 and NB algorithms in the prediction of the student's success in the undergraduate program [13]. The linear random forest (LRF) have advantages over least squared linear regression, neural networks, epsilon support vector regression, KNN regression, regression tree, regression random forest, gradient descent boosted trees, and linear decision tree algorithms in the context of learning ability, algorithm robustness, and feasibility of the hypothesis space [14]. Fuzzy ANFIS was used to convert multiple decision maker's ratings into a final rating based on 78 fuzzy logic [15].

To our best knowledge, prediction of student progress using fuzzy ANFIS and machine learning algorithms is not yet to be done. Using RFC along with fuzzy logic could give a better result and give a new direction toward education data mining.

3. Methods

This research focuses on the identification and analysis of different types of factors including psychological, personal, teaching impact, university facilities, learning environment, etc. which affect students' results and prediction of student's performance based on these factors using machine learning algorithms. Basically, this work can be classified into four major steps as follows:

I. Identify factors through a student survey and an expert survey.
II. Modification of factor rating applying expert's opinion by using fuzzy ANFIS.
III. Classification and prediction of student's performance based on RF classifier.
IV. Analysis of results.

In the 1st step, a survey consisting of 24 questions was done via google response form to get the ratings of copious factors from students. Here the course teachers and student advisors were considered experts. In the following step, factors were reduced by factor analysis and based on selected factors, another survey on experts was conducted. Then the fuzzy ANFIS model was developed to convert multi-response ratings into a single rating point by consisting of fuzzy logic. After that RF was applied using train and test data to classify the student performance. Finally, the results were analyzed with relative advantages and disadvantages.

3.1 Process Flowchart

At first, surveying was done on students to identify the primary factors which affect their performance.

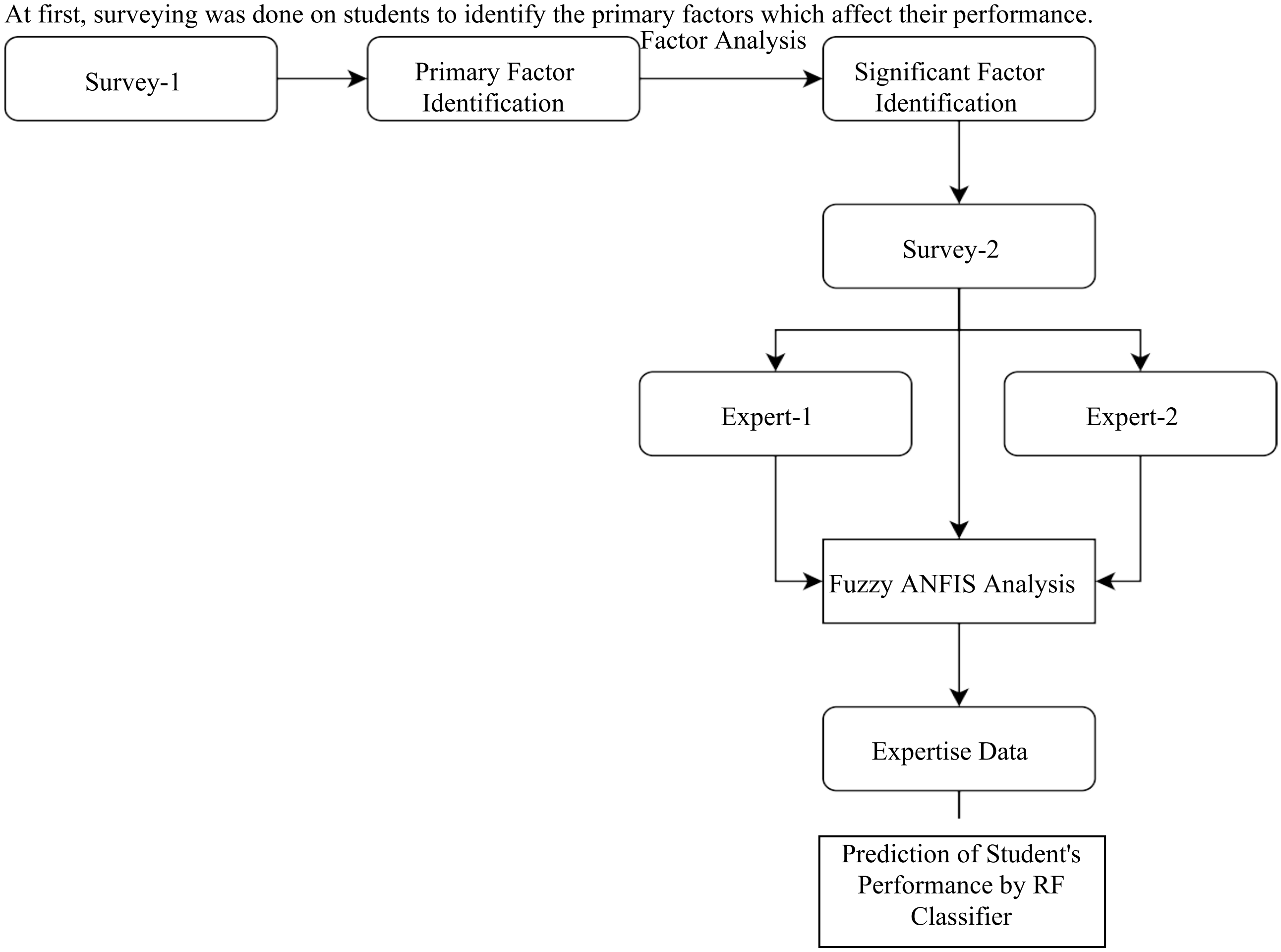

Figure 1: Process flowchart of the proposed model

After performing factor analysis based on students' responses to the survey and experts' opinions, important factors were identified. Then, the second survey was done on experts who gave ratings of the effect of each factor on each student. After the completion of these surveys, ratings of each attribute were merged with the factor ratings given by the experts with the help of Fuzzy ANFIS analysis. Then, 80 percent of these merged data were used as training data for RF classifiers. The other 20 percent was used for testing. After that, the accuracy of the method was checked with the actual results of the students. The full process flowchart of the work is shown in figure 1.

3.2 Fuzzy ANFIS analysis:

Fuzzy logic imitates in a way that resembles human reasoning. It is an approach where computing is based on different "degrees of truth" rather than Boolean (1, 0) logic on which a modern computer is based. A computer can only take precise inputs and give output as TRUE or FALSE which resembles a human's YES or NO. The inventor of fuzzy logic, Dr. Lotfi Zadeh first observed that unlike computers, humans' decision modeling are a range of possibilities between YES and NO, such as [16]-

Table 1: Range of possibilities in fuzzy logic

Certainly Yes	Cannot Say	Possibly Yes	Possibly No	Certainly No

The fuzzy logic works on these possible levels of inputs to give a specific output.
Architecture: Its architecture contains four parts-

- Rule base: It contains IF-THEN rules provided by the experts which help to govern the decision-making system.
- Fuzzification: It is used for converting inputs which are called crisp numbers into fuzzy sets. Crisp inputs are nonetheless the same inputs whose computation is performed by the sensors and it is then passed into the control system for processing.
- Inference Engine: It determines the matching degree of the current fuzzy input with respect to each rule and stimulates human reasoning on the basis of these rules then it decides which rules are to be fired according to the input field. Then, the fired rules are combined to form the control actions.
- Defuzzification: It is used to convert the fuzzy sets obtained by the inference engine into a crisp value which is the ultimate output.

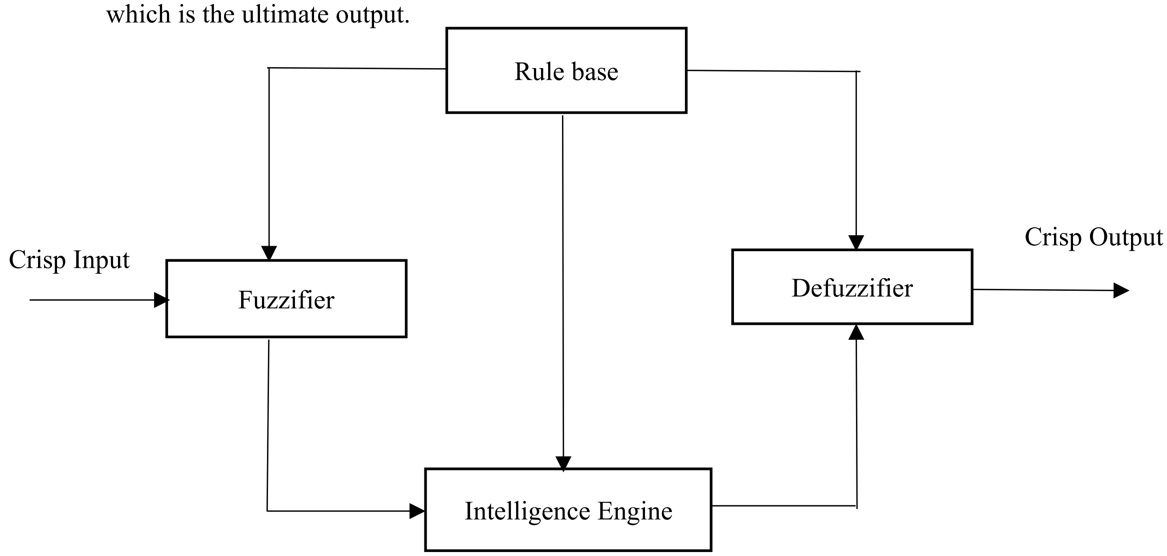

Figure 2: Fuzzy Logic structure

Membership Function: It is a graph that defines how each point in the input and output space is mapped to membership value between 0 and 1. There are largely three types of fuzzifier:

- Singleton fuzzifier
- Gaussian fuzzifier
- Trapezoidal or triangular fuzzifier

In this paper, the fuzzy ANFIS model has been established where a set of inputs and outputs are given then rules are added. At first, the membership functions of each input and output is specified by selecting the type of fuzzifier and adjusting its range. Then rules are added. Train data was transformed through the fuzzy ANFIS model.

3.3 Prediction using Random Forest Classifier

Random Forest (RF) is another supervised machine learning algorithm that is also used for both classification and prediction; however, it is mainly applied for classification applications. Forest means trees and the more the trees the more robust the forest is. In the random forest classification method, this model creates different decision trees based on data samples and when new data points are inserted for its class prediction, each decision tree gives one prediction, and finally, the best solution is selected by voting. For an input vector *(x)*, each decision tree will give a vote. Then, $C_{rf}^{B} = majorityvote\{C_b(x)\}_1^B$ where $C_b(x)$ is the prediction of class on b^{th} random-forest tree and C_{rf}^{B} is the final prediction using the majority vote. The main concept behind this model is simple but a powerful one. It is an ensemble method because many uncorrelated models (trees) working as a community definitely outperforms the working of a single model (tree). The reason for this wonderful effect is that the models protect each other from their errors. Choice of attribute selection and pruning methods are necessary for the design of decision trees. There are many attribute selection methods but the most frequently used attribute selection measures in decision tree induction are gain ration criterion [17] and the Gini Index [18]. RFC uses Gini Index method for its attributes' selection which measures the impurity of an attribute with respect to its classes. For a given training set *P*, selecting a sample case randomly and to predict its class as C_i, the Gini index can be written as-

$$\sum \sum_{j \neq i} (f(C_i, T)/|T|)\big(f(C_j, T)/|T|\big)$$

Here, $(f(C_i, T)/|T|)$ is the probability that a selected case belongs to class C_i.
For generating a prediction model, the RFC needs the definition and insertion of two parameters, the number of classification trees desired and a number of predicting variables that are used in each node to grow the trees. For each node, the best split is done by searching selected features. Thus, RFC consists of *N* decision trees where *N* is user-defined value about the number of trees to be grown. When new data points are to be classified, these are passed down to all those trees and then it chooses its class by maximum votes out of *N* votes.

3.3.1 Framework for RFC Model

Input data with various features and an output attribute with different levels are split into two datasets: training dataset and testing dataset. Then bootstrap aggregating and attribute bagging are developed to form a randomly selected decision tree by minimizing the misclassification rate. Finally, the testing dataset is examined to predict the class. The framework for the RFC model is represented by a flowchart in figure 3.

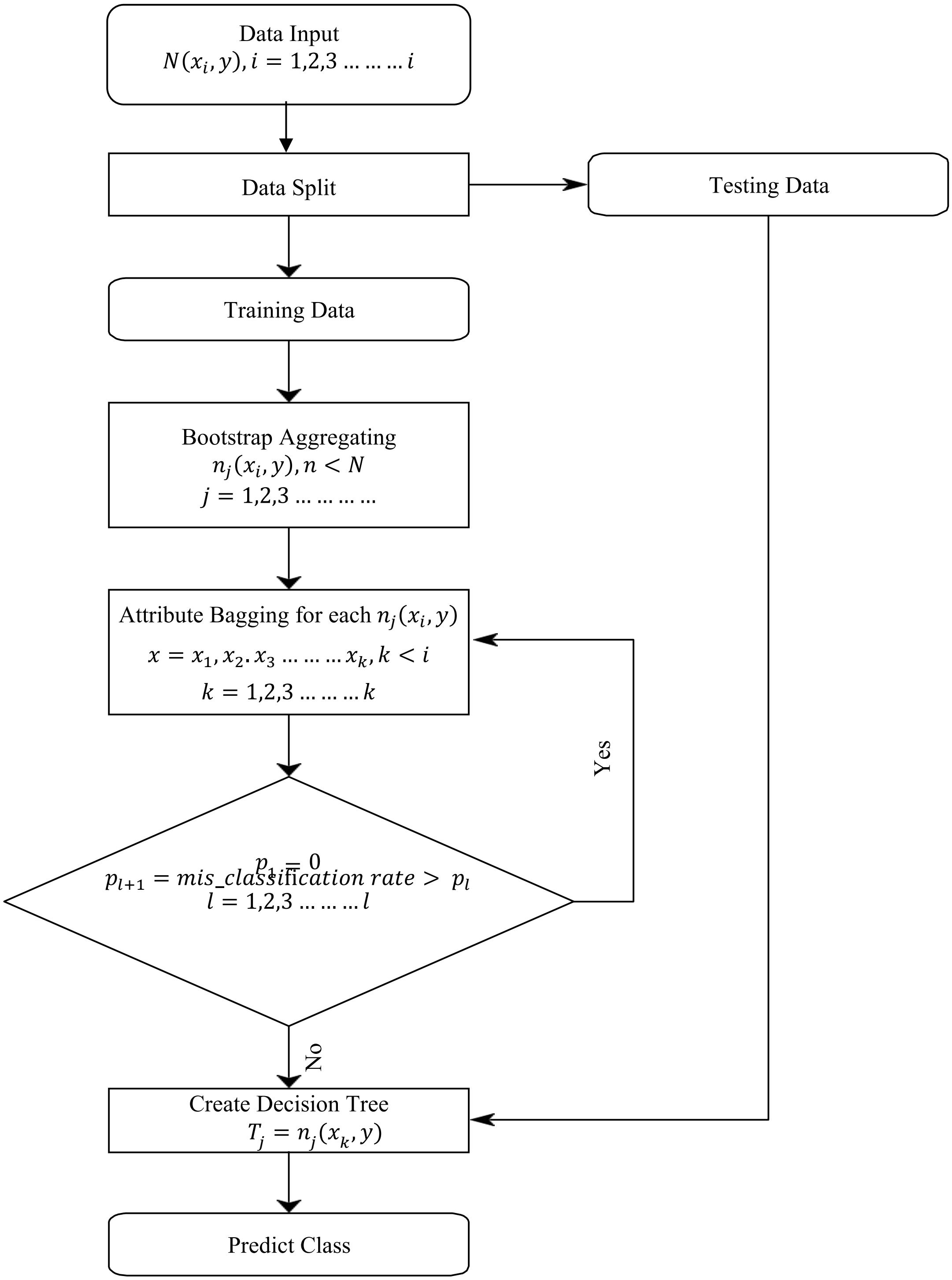

Figure 3: Flowchart of the RFC model

4. Data Collection

In this paper, a survey of 24 questions was first performed on first-year students (45th batch) of the Bangladesh University of Textiles to identify their ratings on each factor which could be linked to their different levels of performance. These 24 factors which were used in survey-1 are shown in Table 2.

Table 2: List of all the factors for survey 1

Creating good notes	Group study	Adaptation to university	Self-confidence
University facilities	Previous year questions	University environment	Effort
Assignment submission	Class tests' marks	Discontent about the university	Motivation
Proper knowledge	Preparation time	Discontent about the department	Procrastination
Exam strategy	Fear of examinations	Course difficulty level	Teaching methods
Residential problems	Hard questions	Family issues	Overconfidence

4.1 Identification of Significant Factors

After performing survey-1, significant factors were identified from the ratings of the students on each factor. Analysis of variance (ANOVA) was used for this purpose to check the correlation between different factors. For the reduction of factors and identification of significant factors, ANOVA had been used here. In Table 3, ANOVA tests' result is shown where two factors, proper knowledge, and time management were compared, and ANOVA with a 95 percent confidence interval was applied. As $F < F-crit$ or $p-value > 0.05$, the null hypothesis is accepted that means the two factors are alike. Another example is given in Table 4, where factors, exam strategy and effort, were compared with the same confidence interval. But result here the shows the opposite. As, $F > F-crit$ or $p-value < 0.05$ that means a null hypothesis is rejected and there is a difference between the factors.

Table 3: ANOVA table of two correlated factors

Factors		Sum of Squares, SS	*df*	*MS*	*F*	*p-value*	*F-crit*	Hypothesis	Result
Proper Knowledge	Between groups	24.49485	1	24.49	0.0511	0.821	3.849	Rejected	Factors Correlated
Motivation	Within groups	557811.5	1164	479.21					
	Total	557836	1165						

Table 4: ANOVA table of two uncorrelated factors

Factors		Sum of Squares, SS	*df*	*MS*	*F*	*p-value*	*F-crit*	Hypothesis	Result
Exam Strategy	Between groups	8215	1	8215.28	17.20	3.6E-05	3.8494	Accepted	Factors uncorrelated
Effort	Within groups	555889	1164	477.56					
	Total	564104	1165						

In this way, using the ANOVA tests, correlated and uncorrelated factors were identified. Then correlated factors were eliminated and uncorrelated factors were taken as significant factors (factors responsible for students' different levels of performances) which were used for the whole classification problem. 11 significant factors were found which are shown in Table 5.

Table 5: List of 11 significant factors

Factors' no.	Factors' name	Factors' no.	Factors' name
1	Creating good notes	7	Political Involvement
2	Exam strategy	8	Personal problems
3	Class tests marks	9	Fear of examinations
4	Teaching methods	10	Procrastinations
5	Hard questions	11	Effort
6	Adaptation to university		

4.2 Merging of factors' ratings by Fuzzy analysis

After finding 11 significant factors using ANOVA analysis, another survey was conducted on two experts (one expert was course teacher and another expert was course coordinator) with those factors. Now, there were three ratings for each factor and fuzzy was used to merge these ratings. To merge these ratings in fuzzy analysis, different rules were imposed on inputs (students' and experts' ratings) to get the possible output (merged value). Figure 4 shows the demonstrations of these rules on inputs and how they affect the outputs.

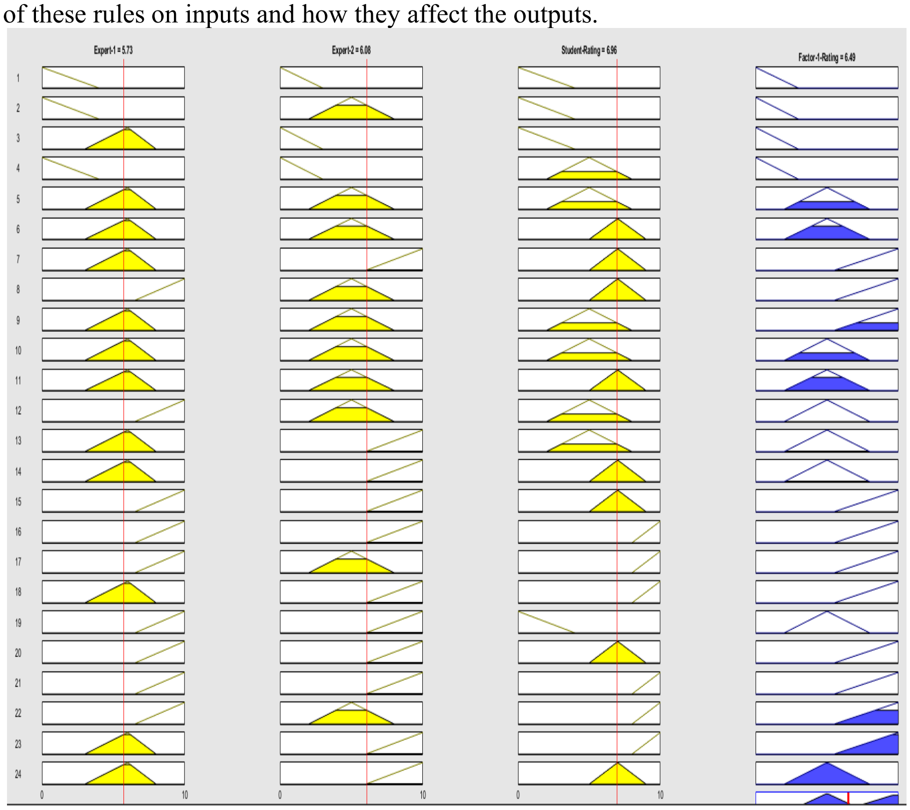

Figure 4: Demonstration of rules on fuzzy ANFIS

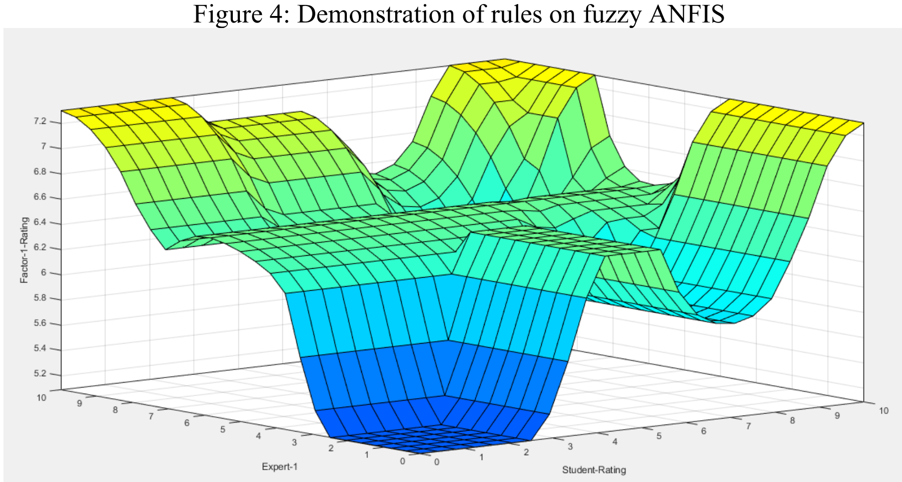

Figure 5: Factor rating with respect to expert-1 vs student rating.

How to factor rating is affected by any two input ratings are shown in Figure 5. After that, combining all these outputs, finally, the merged ratings were identified.

5. Results and Discussion

Random forest was created using a training dataset where 500 trees were built. In Figure 6, one of the random trees is presented.

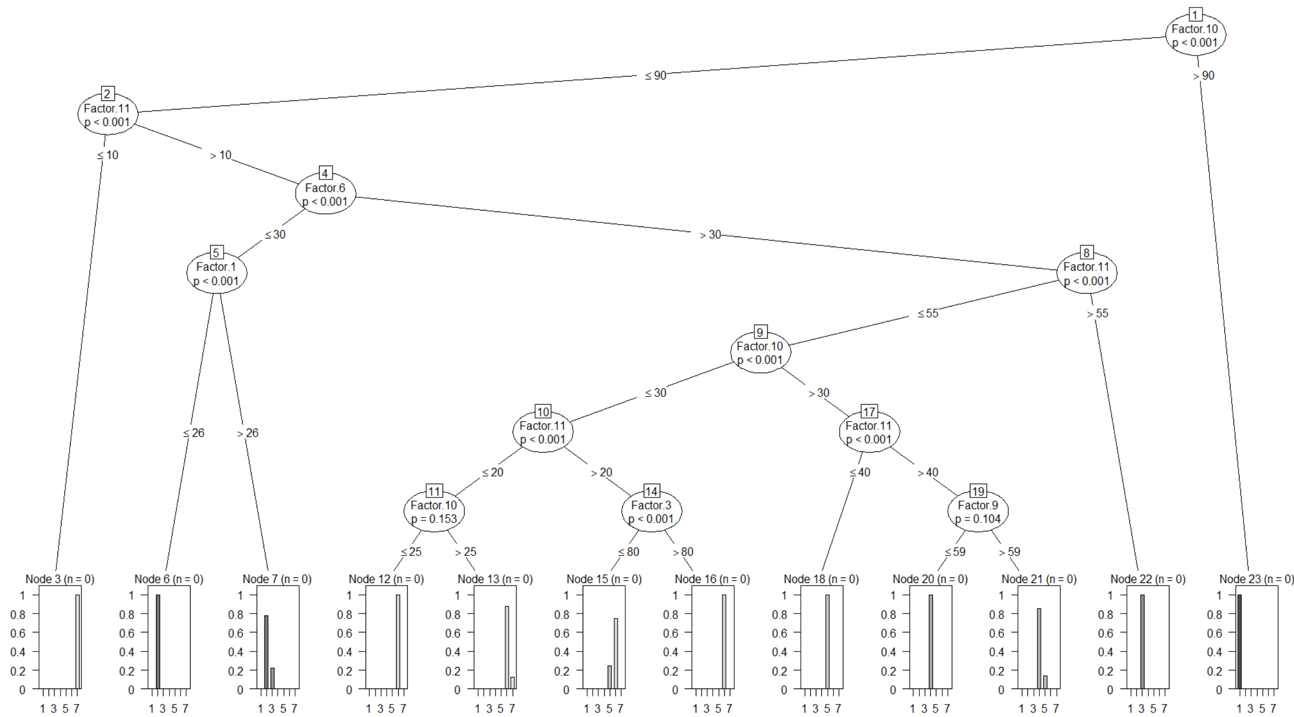

Figure 6: One of the random forest trees with probable outcomes

After that error rate was calculated which is shown in the figure. We can see that above 150 trees, the error rate became constant. That is why 150 trees were selected to build the RF model. The error rate for 150 trees is also given in Figure 7.

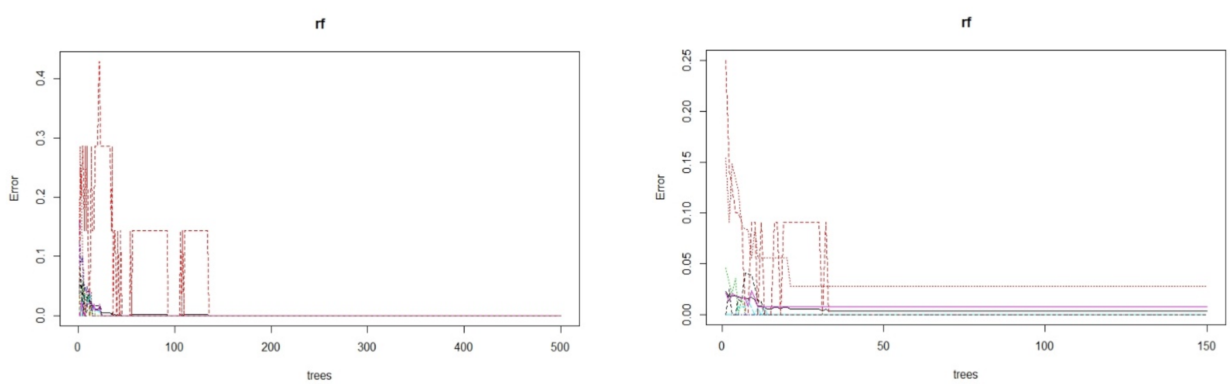

Figure 7: Error rate with respect to no of trees

Another important parameter of RF classifier is the number of attributes used in the attribute bagging process which is m_{try}. Analyzing the OEB error to find out the suitable m_{try}. The relationship between OEB error and m_{try} is depicted in Figure 8. The lowest OEB error is found when the value of m_{try} is less than 3.

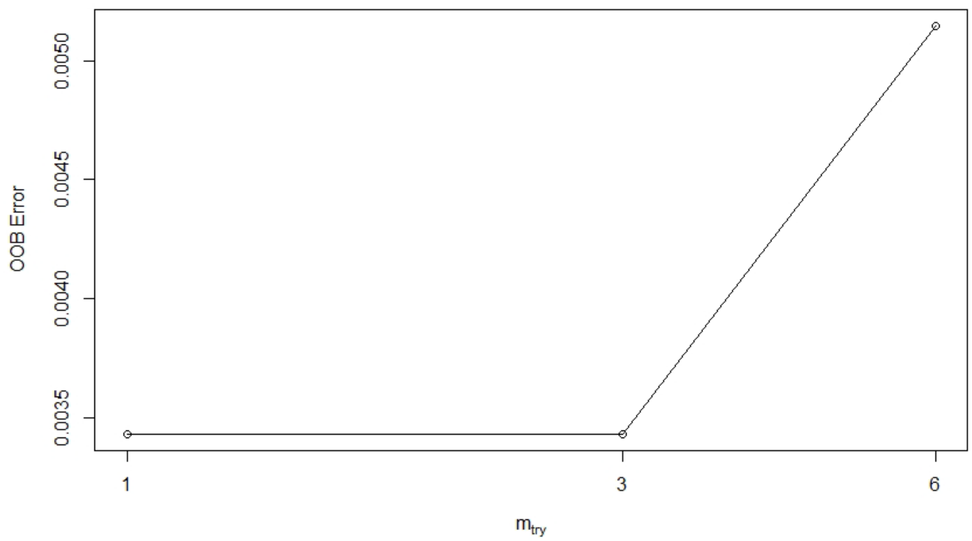

Figure 8: OEB error vs m_{try} graph

The partial dependency of individual factors on different classes was also investigated. For example, the partial dependency of factor-1 for two distinct classes is shown in Figure 9. It depicts that if the factor value is between 60 to 80 it gives a more accurate value for predicting class 5 while it predicts more accurately the class 3 when it is below 45.

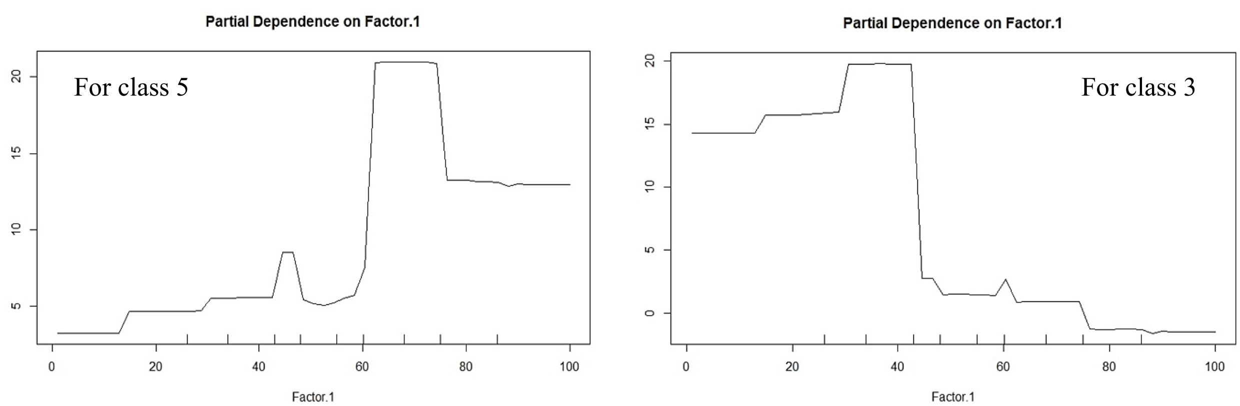

Figure 9: Partial dependency on factor-1

The accuracy of the RF classifier is 96.88%. The confusion matrix for this model is shown in Table-6. This model can not predict class 6 accurately.

Table 6: Confusion matrix for the RFC model

		Predicted Class by RFC						
		1	2	3	4	5	6	7
Actual Class	1	1	0	0	0	0	0	0
	2	0	6	0	0	0	0	0
	3	0	0	16	0	0	1	0
	4	0	0	0	4	0	0	0
	5	0	0	0	0	3	0	0
	6	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	1

The important factors were also analyzed using RFC which is depicted in Figure 10. We can see the top 5 important variables. The 1st graph tests how worse the model performs without each variable.

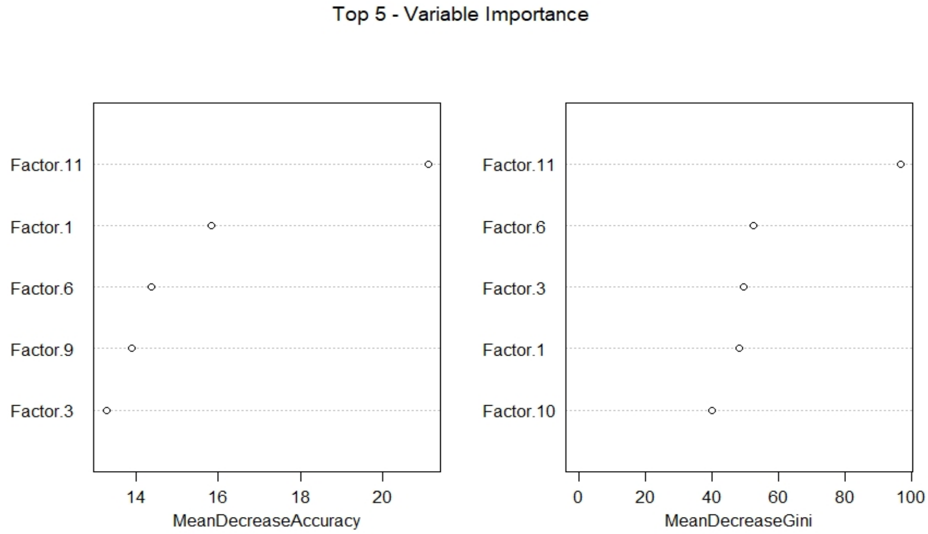

Figure 10: Top five factors that affect student performance most

Factor-11 (effort of the student) is the most important factor that is responsible for better prediction of student performance. In the 2nd graph, it measures how pure the nodes are at the end of the tree without each variable. For factor-11, 6, 3, 1, & 10 mean decrease in Gini is very high. Thus, these factors are responsible for the student's performance. Authority needs to focus on those factors so that the students' performance can enhance.

6. Conclusion

In this work, we propose a hybrid model for predicting first-year university students' performance in the final examinations. The fuzzy ANFIS is incorporated with RFC to develop that model. Our experimental results illustrated that our proposed model is proved to be effective and pragmatic for the accurate prediction of students' progress, as compared to some traditional machine learning algorithms. The accuracy of the prediction of the RF model is 96.88%. It is also found that the students' effort and creating good notes (factor-11 and factor-1) are the two key factors that have a great impact on students' success. So, early detection of these factors' ratings could yield valuable insights for a better educational environment and assistance to students' performance.

In conclusion, we point out that the students' attributes implemented in our work are not bounded rather more new attributes can be introduced in our database to improve the quality of our model. Not only new attributes but also more experts can be added to get more insight into factor ratings. This proposed model can be applied to analyze the senior students' performance also. Besides, support vector machine (SVM), deep learning algorithms like Convolutional Neural Network (CNN), Recurrent Neural Networks (RNNs) can be implemented for comparative analysis. However, more data could be integrated by considering other university students' conditions, which would be more versatile. In addition, students' progress can be evaluated subject-wise.

References

[1] T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using Educational Data Mining," Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016, pp. 91–95, 2016, doi: 10.1109/SAPIENCE.2016.7684167.

[2] C. Anuradha and T. Velmurugan, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance," Indian Journal of Science and Technology, vol. 8, no. 15, 2015, doi: 10.11591/ijece.v8i5.pp3966-3975.

[3] A. Acharya and D. Sinha, "Early Prediction of Students Performance using Machine Learning Techniques," International Journal of Computer Applications, vol. 107, no. 1, pp. 37–43, 2014, doi: 10.5120/18717-9939.

[4] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques," Applied Mathematical Sciences, vol. 9, no. 129, pp. 6415–6426, 2015, doi: 10.12988/ams.2015.53289.

[5] P. A. Patil and R. v. Mane, "Prediction of Students Performance Using Frequent Pattern Tree," in Proceedings of the 6th International Conference on Computational Intelligence and Communication Networks, 2014, pp. 1078–1082, doi: 10.1109/CICN.2014.226.
[6] I. E. Livieris, V. Tampakas, N. Kiriakidou, T. Mikropoulos, and P. Pintelas, "Forecasting Students' Performance Using an Ensemble SSL Algorithm," in International Conference on Technology and Innovation in Learning, Teaching and Education, 2018, pp. 566–581, doi: 10.1007/978-3-030-20954-4_43.
[7] F. Okubo, A. Shimada, T. Yamashita, and H. Ogata, "A Neural Network Approach for Students' Performance Prediction," in Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 2017, pp. 598–599, doi: 10.1145/3027385.3029479.
[8] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," SN Applied Sciences, vol. 1, no. 9, 2019, doi: 10.1007/s42452-019-0884-7.
[9] V. Tampakas, I. E. Livieris, E. Pintelas, N. Karacapilidis, and P. Pintelas, "Prediction of students' Graduation time using a two-level classification algorithm," in International Conference on Technology and Innovation in Learning, Teaching and Education, 2018, pp. 553–565, doi: 10.1007/978-3-030-20954-4_42.
[10] M. Chauhan and V. Gupta, "Comparative Study of Techniques Used in Prediction of Student Performance," World Scientific News, vol. 113, pp. 185–193, 2018.
[11] B. Jia et al., "Prediction for Student Academic Performance Using SMNaive Bayes Model," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11888 LNAI, pp. 712–725, 2019, doi: 10.1007/978-3-030-35231-8_52.
[12] Y. Abubakar and N. B. H. Ahmad, "Prediction of Students ' Performance in E- Learning Environment Using Random Forest," International Journal of Innovative Computing, vol. 7, no. 2, pp. 1–5, 2017.
[13] T. Mahboob, S. Irfan, and A. Karamat, "A machine learning approach for student assessment in E-learning using Quinlan's C4.5, Naive Bayes and Random Forest algorithms," 2017, doi: 10.1109/INMIC.2016.7840094.
[14] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling," Journal of Petroleum Science and Engineering, vol. 174, pp. 776–789, 2018, doi: 10.1016/j.petrol.2018.11.067.
[15] S. K. Ghosh, N. Zoha, and F. Sarwar, "A Generic MCDM Model for Supplier Selection for Multiple Decision Makers Using Fuzzy TOPSIS," in Proceedings of the 5th International Conference on Engineering Research, Innovation and Education (ICERIE) Sylhet, Bangladesh, 2019, pp. 833–840.
[16] L. A. Zadeh, "Fuzzy Logic," Computer, vol. 21, no. 4, pp. 83–93, 1988.
[17] S. Salzberg, "Book review: C4. 5: by j. ross quinlan. inc., 1993. programs for machine learning morgan kaufmann publishers," Machine Learning, vol. 16, pp. 235–240, 1994, doi: 10.1016/S0019-9958(64)90259-1.
[18] C. Liu, M. White, and G. Newell, "Measuring the accuracy of species distribution models: A review," in Proceedings 18th World IMACs/MODSIM Congress. Cairns, Australia, 2009, pp. 4241–4247, [Online]. Available: http://mssanz.org.au/modsim09.

Biography

Sourav Kumar Ghosh is a lecturer in Industrial and Production Engineering at Bangladesh University of Textiles (BUTEX). He earned B.Sc. in Industrial and Production Engineering from Bangladesh University of Engineering and Technology (BUET) in 2017. He is a former lecturer in textile engineering at Primeasia University. He is currently enrolled in a Master's program in Industrial and Production Engineering at Bangladesh University of Engineering and Technology (BUET). He has published six journal papers and eight conference papers. S. K. Ghosh has completed several research projects under UGC. His research interests include machine learning, supply chain optimization, operation research, parameter optimization of CNC machines, renewable energy, and lean manufacturing.

Farhatul Janan is a Lecturer in the Department of Industrial and Production Engineering at Bangladesh University of Textiles, Dhaka, Bangladesh. She previously worked as a Lecturer in the Department of Industrial and Production Engineering at Military Institute of Science and Technology, Dhaka, Bangladesh. She earned B.Sc. in Industrial and Production Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh. She is an ongoing student of M.Sc. in Industrial and Production Engineering in Bangladesh University of Engineering and Technology. Her research interests include operations research, machine learning and regression analysis, data science, supply chain management and logistics systems optimization, engineering economics and decision theory.