IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 9, No. 1, March 2020, pp. xx~xx
ISSN: 2252-8938, DOI: 10.11591/ijai.v9.i1.ppxx-xx

MS and PhD Admission Prediction of Bangladeshi Students into Different Classes of Universities

Md Naimul Islam Suvon[1], Sadman Chowdhury Siam[2], Mehebuba Ferdous[3], Mahabub Alam[4] Riasat Khan[5]
[1,2,3,4,5]Department of Electrical and Computer Engineering, North South University, Bangladesh

Article Info

Article history:

Keywords:

random forest
decision tree
bootstrap aggregating
data normalization
on-hot encoding

ABSTRACT

Every year, many Bangladeshi students intend to pursue higher studies abroad after completing their undergraduate degrees. Choosing a university for higher education is an ambiguous task for students. Usually, they face various problems to select the perfect university for them according to their profile. Especially, the students with average and lower academic credentials (undergraduate grades, English proficiency test scores, job, and research experiences) can hardly choose the universities that could match their profile. In this paper, we have analyzed some real unique data of Bangladeshi students who had been accepted admissions at different universities worldwide for higher studies. Finally, we have produced prediction models, which can predict appropriate universities of specific classes for students according to their past academic performances. Two separate models have been studied in this paper, one for MS students and another for PhD students. The universities where the students got admitted have been divided into four classes according to the QS World University Rankings. Random Forest and Decision tree algorithms are used for making the multi-class classification models. F1-score, accuracy, weighted accuracy, and the receiver operating characteristic curves have been studied for the three machine learning algorithms. Numerical results show that the random forest algorithm demonstrates the highest accuracy for both the PhD and MS data with approximately 70% accuracy.

Corresponding Author:

Md Naimul Islam Suvon
Department of Electrical and Computer Engineering,
North South University University,
Shaymoli-25, Dhaka, 1207, Bangladesh
Email: naimul.suvon@northsouth.edu

1. INTRODUCTION

Nowadays, educational data has become more popular among researchers. Educational information mining is the process of acquiring necessary information from an extensive collection of educational datasets and finally making significant decisions from them [1]. Many students of Bangladesh apply for higher studies every year in different universities all over the world. The students spend a significant amount of money and time to prepare for the application process. Unfortunately, most of them face difficulties deciding which universities they should apply to according to their different test scores. Many students tend to choose the safe options, where there are high possibilities to get admitted. Conversely, some of them apply to an ambitious higher-class university, which does not conform with their academic profile, leading to an ultimate rejection. Many students face this kind of problem as they cannot evaluate their academic credentials according to the admissions criteria. There are plenty of consultancy centers in Bangladesh, where they evaluate the students' profiles and provide guidance for the application process in exchange for high consultation fees. But eventually, they failed to find the perfect match of universities for students to apply; because they cannot evaluate their profile correctly. Sometimes they are misled by the senior graduates about the university's ranking and past admission decision patterns.

In this paper, we have worked to make the university selection procedure easier for the students according to their academic profile. We work on a machine learning-based approach to predict the perfect university match for students pertaining to their past academic records, i.e., undergraduate university and CGPA, English proficiency test’s score, job experience, and research paper. The students can evaluate their chance of getting admission to a higher rank or lower rank university. This paper's primary contribution is to work on an exclusive real dataset of Bangladeshi undergraduate students who have gone for higher studies abroad to USA, Canada, Germany, Australia, and different foreign countries for MS and PhD degrees for the last two years of 2018 and 2019 from more than 30 universities of Bangladesh. The initial dataset contains a lot of features, including the undergraduate university, subject and grades, admitted university, subject and research area, funding sources, GRE and English aptitude test scores, research papers, job or research experiences, etc. Next, we have analyzed this data to find the essential features that we need for our model. We have divided the accepted universities into four classes according to this year's QS World University Rankings [2]. Subsequently, we have developed three different approaches (each for the MS and PhD students' data) to make the model for assessing the possibility of a student's admission to a particular class of university. We have implemented the decision tree algorithm [3] and two ensemble learning methods, random forest [3], and adaptive boosting [4] in this work. Finally, we have reported all the machine learning algorithms' performance for both the MS and PhD applicants' data in terms of the evaluation metrics, e.g., precision, recall, F1-score, accuracy, weighted accuracy, ROC curve, AUC, [5] etc. To the best of our knowledge, this is the first time various multi-class classification models to select different universities worldwide have been done on the dataset of Bangladesh's students.

This paper is categorized as follows: Section 2 discusses the related work, and the proposed system methodology has been described in Section 3. In Section 4, we have shown the performance evaluation of our system for the implemented algorithms with appropriates tables and figures. Finally, Section 5 concludes the following work with some possible directions in future research.

2. LITERATUE REVIEW

In this section, related papers that are similar to our work have been discussed. In the paper [6], N. T. N. Hien and P. Haddawy used the Bayesian network's approach to predict the graduating student's cumulative grade point average based on the applicant's background (previously attended institutions, undergraduate CGPA, English test score, the field of study, age, gender, marital status, etc.) at the time of admission. They evaluated the stratified ten-fold cross-validation technique of three years' admissions data of the Asian Institute of Technology (AIT), Thailand. The study shows a mean absolute error of 0.22-grade points for a master's program and 0.20 for the doctoral program. The study used the Bayesian network's approach; it gives departmental faculty members valuable information in making admission decisions. From the Bayesian network prediction model, they represented a case-based retrieval mechanism that the same similarity measure used by the case-based system. The case-based system shows the past student most similar to the evaluating students.

In this paper [7], A. Waters and R. Miikkulainen estimated the chance of admission of new applicants based on past admissions decisions at the Department of Computer Science of University of Texas at Austin, USA. They used a statistical machine learning technique (L1 regularized logistic regression) to evaluate this system from different numerical, categorical, and text features data. This system predicts a real-valued score for every student's file, similar to the traditional human reviewers. The proposed system GRADE (graduate admissions evaluator) attained an accuracy of 87.1% and reduced the total review time by 74%.

In this paper [8], M. S. Acharya, A. Armaan, and A. S. Antony used Machine Learning based methods, where they compared different regression algorithms to predict the applicants' chance of graduate admissions. They used linear regression to predict results and support vector regression to use kernel trick to predict data. Next, they implemented a decision tree that breaks the dataset sequentially into a smaller subset, and in the meantime, the associated decision tree was developing accordingly. Finally, they applied random forest regression. It is an addictive type model, and this model helps to predict by combining decision from a sequence of base models. Each base model is a decision tree, and the result of the random forest model is the cumulative output of the decision trees. They used multiple models to get excellent predictive work, which is known as model assembling. The authors found that the linear regression achieved the highest accuracy on their dataset (hypothetical open-source data of UCLA), which had low MSE and a high R2 score compared to the other implemented regression techniques.

In this paper [9], I. Hmiedi et al. made a regression model using the Random Forest Algorithm to predict the graduate admissions probabilities. This work used the same hypothetical open-source dataset from

Kaggle of the University of California in Los Angeles as in [8]. The authors applied data augmentation to achieve a more diverse dataset and reduce overfitting and data preprocessing (data normalization and duplicate removal). They split the data into 70% for training and 30% for the testing and finally reported the proposed model's accuracy.

P. Janani et al. predicted the chance of graduate admissions using the decision tree Algorithm in [10]. They used a classification algorithm with the decision tree classifier to predict the output due to its simple logic, effectiveness, and interpretability. This model works by creating a tree-like structure by dividing the dataset into several smaller subsets based on different conditional logic. The authors attained 93% accuracy by using the same open-source dataset in [8] and the decision tree classifier in output

3. METHODOLOGY

Our approach for making the model is divided into different sections. Figure 1 represents a flowchart of our proposed model. Here RF, DT, and AdaBoost mean random forest, decision tree, and adaptive boosting, respectively. These are different classification algorithms that we will use in our model. In the subsequent sections, the working methods of this paper have been described in detail.

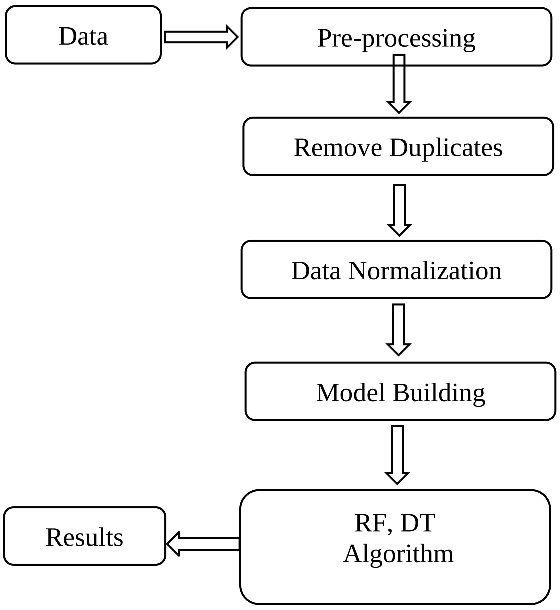

Figure 1. Proposed System Model.

3.1. Data Collection

The primary contribution of this work is to create a unique dataset. We have collected our dataset from the Graduate Resources Enhancing Center (GREC), Bangladesh. GREC is one of the largest platforms for Bangladeshi students, where every year, many students take preparation for various standardized exams, e.g., GRE, IELTS, TOEFL, SAT. They collect information from the students who are admitted to different universities worldwide with scholarships. We have collected the data for the last two years, 2018 and 2019, for students who have got acceptance in various universities in mostly USA, Canada, Australia, Germany, and UK. Initially, there were 230 students' data from more than 30 universities of Bangladesh, which contains masters, as well as doctoral applicants. Also, a single candidate got a chance in multiple universities simultaneously. As expected, the academic credentials are scattered quite differently between the master and doctoral students, with doctoral students were tending to achieve better academic performance than the master's students. Next, we separated the MS and PhD data and made two datasets to apply different prediction models for them. Then we have expanded our dataset by processing it into the multiple universities accepted candidate's data. Finally, we obtained approximately 400 data for PhD students and 300 for the MS candidates. As this total 700 data has been obtained from 230 candidates, an individual student got accepted in average 3 universities. Next, we have added a new feature to our dataset, the universities' QS World University Rankings, where the students have been admitted. QS World University Rankings, partnered with Elsevier, is the most accepted international rankings of universities worldwide. According to the university rankings, we have divided the candidates into four classes, where they have been admitted. Class A is the candidates who have been accepted in a university with a QS World University Rankings between 1 to 250. Similarly, classes B, C, and D are for university rankings between 251 to 500, 501 to 750, and above 750, respectively. Our work's final objective is to find in which class of university a candidate should apply with his academic profile.

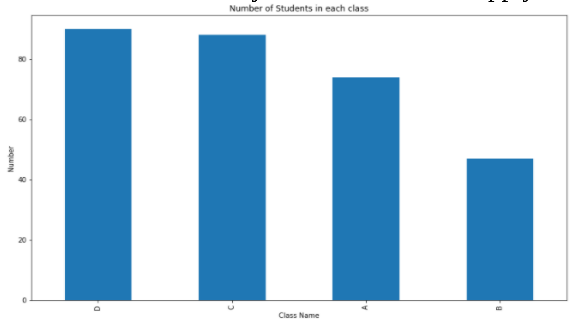

Figure 2. Number of MS students in each class of university.

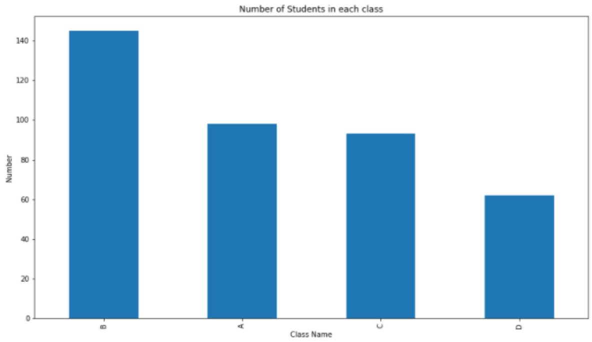

Figure 3. Number of PhD students in each class of university

Figure 2 and Figure 3 show the number of MS and PhD students in each university class. Our initial dataset contains many features in three forms, numerical, categorical, and text features. Those are student's name, admitted university name with its state and country, admitted department, intended research area, types of funding (fellowship, assistantship, external scholarship), intended semester, undergraduate university name with CGPA and department, IELTS/TOFEL score, GRE score, publications (conference or journal), job experience, research experience, application method, funding source, etc.

3.2. Data Analysis

It is crucial to analyze data before building a model because there might be many missing, inconsistent, and duplicate data [11]. We have performed some analysis on both the MS and PhD dataset so that the applied algorithm can quickly analyze it. Most of the students' data are structured in a uniform format so that the algorithm can easily interpret it, except the undergraduate and admitted universities' names. Some students may describe the same undergraduate institution as BUET, Bangladesh University of Engineering and Technology, Bangladesh U of Engineering and Technology, etc. The previously attended undergraduate universities' names are rephrased in uniform abbreviated string formats at the first data preprocessing step, e.g., BUET, DU, RUET, SUST, etc. The admitted graduate universities' names are not required to be preprocessed as they are divided into different classes and consequently have been removed from the feature vectors.

In Figure 4 and Figure 5, we have analyzed the Bangladeshi institutions from where most students got accepted for higher studies for the last two years, 2018 and 2019. For the MS program, the University of Dhaka (DU), and Bangladesh University of Engineering and Technology (BUET), have the highest number of students accepted into different universities worldwide. On the other hand, BUET is at the top, and DU is the second position for the PhD program. Not surprisingly, both of these universities' students are mostly dominating in both programs. They are the best undergraduate institutions in Bangladesh regarding academic and employer reputation, acceptance rate, tuition fees, faculty-to-student ratio, etc.

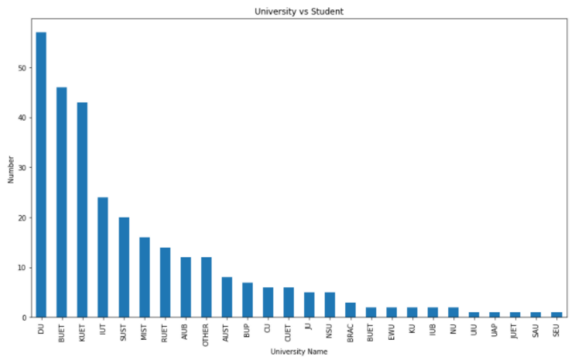

Figure 4. The number of students who got accepted for MS according to the undergraduate university.

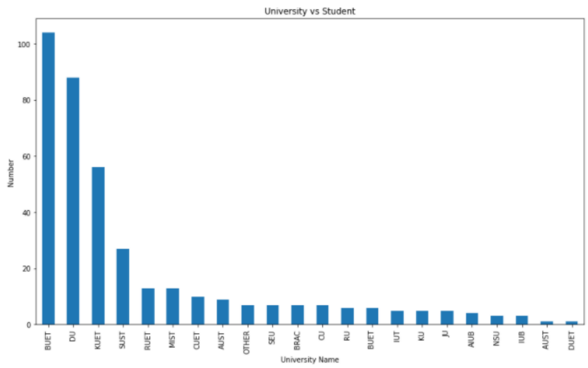

Figure 5. The number of students who got accepted for PhD according to undergraduate university.

Next, we have found out the important features [12] which mostly influenced our model. Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. The higher node probability value refers to the more important feature. The final feature importance, at the random forest level which we have used, is its average over all the trees. The sum of the feature's importance value on each tree is calculated and divided by the total number of trees, as shown in Equation 1.

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T} \tag{1}$$

In Equation 1, $RFfi_i$ is the importance of feature i calculated from all the tress in the random forest model, $normfi_{ij}$ denotes the normalized feature importance for i in tree j and T is the total number of trees.

$$normfi_{ij} = \frac{fi_i}{\sum_{j \in all\ features} fi_j} \tag{2}$$

In Equation 2, each feature's importance on a decision tree is normalized to a value between 0 and 1 by dividing by the sum of all the feature importance values. Here fi_i is the importance of feature i, which can be calculated by Equation 3. In Equation 3, ni_j is the importance of node j, which can be obtained from Equation 4 using Gini importance, assuming only two child nodes (binary tree). Here w_j is the weighted number of samples reaching node j, C_j. The impurity value of node j and $left(j)$ and $right(j)$ is the child node from left and right split on node j, respectively.

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k} \tag{3}$$

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{4}$$

In Figure 6, for MS data, it can be seen that the essential feature is undergraduate CGPA, which score is almost 0.7, and the second one is GRE, with a score of 0.15. Similarly, in Figure 7, for Ph.D. data, Bachelor's CGPA and GRE are the top two needed features for our model. So, it means that for our prediction, these two features will perform a crucial role.

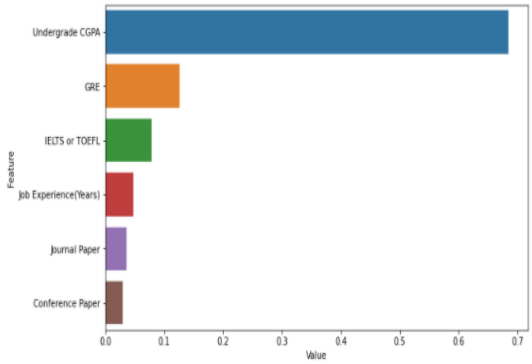

Figure 6. Feature importance scores of MS students' data.

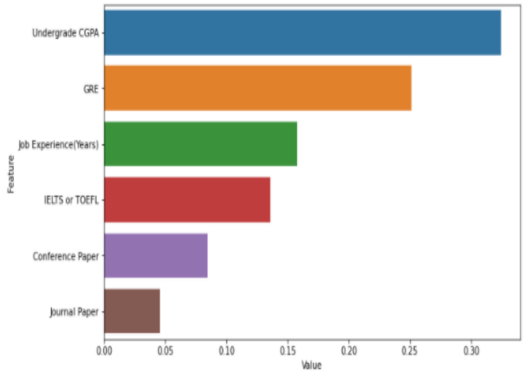

Figure 7. Feature importance scores of PhD students' data

In our raw data, there was a lot of information about an individual student. Our model returned significantly biased results with all the provided data with a low accuracy score on the training step. We performed feature engineering and dimensionality reduction of the raw data and selected the following eight features, which consequently improves the predictive model's performance and effectiveness.

1. Undergrad University
2. Undergrad CGPA
3. TOEFL/IELTS
4. GRE
5. Conference Paper
6. Journal Paper
7. Job Experience (Years)
8. Undergrad University Class

Then we applied the required preprocessing to these data features, which are described in the following section.

3.3. Data Preprocessing

In this step, we have gone through a few steps to perform data cleaning and data transformation . Firstly, we have removed missing data from the table. Since we worked with GRE, IELTS, and TOEFL test results, we found some of the test results were unavailable as GRE is only required at the United States universities, and all the universities require either IELTS or TOEFL. We normalized the test scores and merged them into a single feature using Equation 5. We also one-hot encoded the categorical data by creating binary or dummy variables for using them as input in the machine learning model. We then applied a transformation to numerical data by performing normalization on them.

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{5}$$

For our prediction, we have classified the universities based on the ranking into four classes (A, B, C, D) as explained in Section 3.1.

3.4. Data Preprocessing
3.4.1. Decision Tree
We decided to use Decision trees as they are statistical models that can generate trees based on each feature's information gain and return the predicted output by making decisions based on the tree nodes [3]. We chose the ID3 algorithm [13], which is mostly used to generate decision trees. It uses a greedy top-down approach to choose the tree nodes base on the information gain of a particular feature, and Equation 6 is used to measure the information gain. We calculated the gain by the difference in the parent node's entropy and its corresponding child nodes' average entropy. Here IG is information gain, WA denotes weight average, $E(P)$and $E(C)$ represent parent and child entropy respectively.

$$IG = E(P) - WA \times E(C) \tag{6}$$

Equation 7 is used for measuring the entropy, which is a measure of disorder obtained by selecting a particular feature.

$$E(S) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{7}$$

The ID3 algorithm constructs the trees by choosing a variable with maximum gain and splits the data based on the attributes (if attribute is available or not in the data). It keeps constructing child nodes by selecting different variables until all the features are explored and reaching a conclusive decision.

As our dataset has distinct features like CGPA, standardized test scores, number of publications, etc., the decision tree model is a perfect choice in our case as it can make step by step decisions from each of the features.

3.4.2. Random Forest
We also tried the random forest model, which is an approach based on the ensemble method. It is one of the popular approaches which offers an optimized predictive algorithm by combining several models [3]. In the ensemble method, several weak learners are combined to build a more robust model. We chose decision trees as weak models for our model and then used the weak models to construct a powerful learner. Each weaker model creates a new dataset by considering a random subset of data from every tree's original dataset. Then the tree is trained on the unique subset of data. In this way, we trained a large number of individual decision tree models. The data subset can be chosen with no replacement, i.e., pasting or relief, also known as Bagging or Bootstrap aggregating. The random forest model then follows the prediction that matches the most trees' output like a voting system. We have used random forest in our model as decision trees tend to overfit a lot. So, with random forest, we can eliminate the disadvantage of overfitting by averaging the result.

4. PERFORMANCE EVALUATION
This work's primary objective is to make a reliable prediction of a student's admission into a specific university class. We have analyzed our acquired unique dataset of 400 PhD students and 300 MS candidates with three machine learning methods. At first, we have split the dataset into training and test subset by the ratio of 70:30. For showing the performance of each method, we have found out different evaluation matrices. They are precision, recall, F1-score, accuracy, weighted accuracy, and confusion matrices. Precision indicates the percentage of relevant cases among the retrieved ones (TP divided by $TP + FP$), and recall specifies the percentage of relevant data that have been actually retrieved (TP divided by $TP + FN$). TP and FP means True positive and False Positive respectively, and FN means False Negative. We have also shown the ROC curve for the best-performed model.

4.1. Performance of Decision Tree
The decision tree is one of the most powerful machine learning algorithms that split a node by searching for the most important feature. We have made two models, one for the MS students and another for the PhD students. From Table 1 it can be found that, the accuracy of the MS model is lower than the PhD model. The result of all the other metrics in Table 1 is between 63 to 70 percent. The higher amount of PhD data provides better precision and recall value than the MS data.

According to Figure 8, both models (MS on the left side and PhD on the right side of the figure) have the highest accuracy in predicting the class A university students. The rest of the classes have accuracies in the acceptable range. For class B and C, the overall accuracy of the model is decreasing. There might be too much noisy data in class B and class C students. The class D for the PhD model has the smallest size (60 PhD students of total 400) with the lowest accuracy.

Table 1. F1-Score, Precision, Recall, Accuracy and Weighted Accuracy for Decision Tree

	F1-Score	Precision	Recall	Accuracy	Weighted Accuracy
MS	63.8%	65.7%	63.1%	64%	68.5%
PhD	70.8%	74.4%	67.7%	69%	67.5%

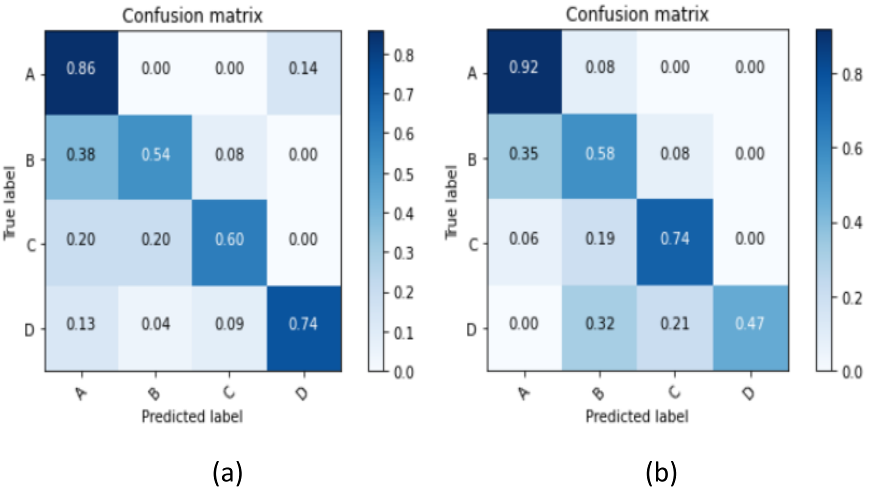

Figure 8. Confusion matrix of the Decision Tree model for MS (a) and PhD (b).

4.2. Performance of Random Forest

We have shown different result metrics of our random forest model in Table 2. It can be seen from Table 2 that we have similar accuracy for both the models, with higher weighted accuracy for the PhD model. The accuracy of the random forest PhD model is similar to its building block decision tree PhD model. Both of them have an accuracy of 69%.

Table 2. F1-Score, Precision, Recall, Accuracy and Weighted Accuracy for Random Forest

	F1-Score	Precision	Recall	Accuracy	Weighted Accuracy
MS	67.1%	64.1%	68.2%	69%	65.25%
PhD	75.1%	71.9%	69.2%	69%	72%

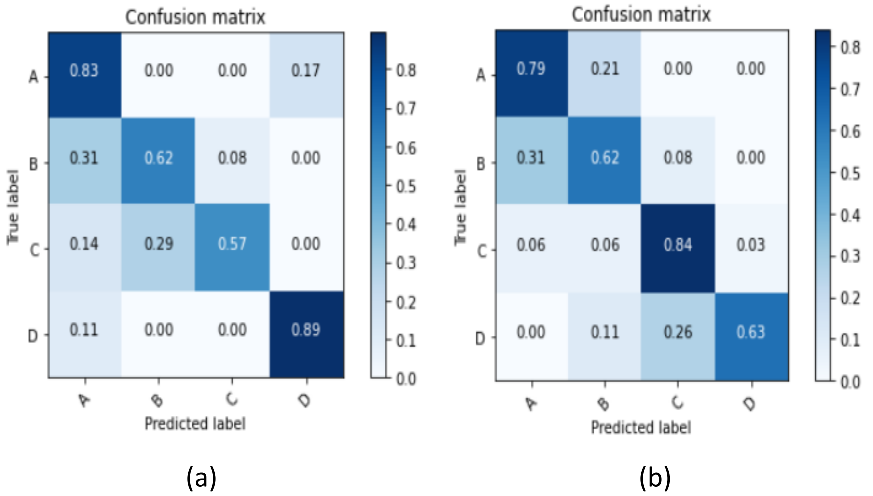

Figure 11. Confusion matrix of random forest model for MS (a) and PhD (b).

In Figure 11, we have shown confusion matrices of our random forest models for MS and PhD data, where weighted accuracy is computed by taking the average over all the classes. From this confusion matrices, we have found out the weighted accuracy for our models. The MS model confusion matrices have higher accuracy in detecting class A and Class D. However, the total accuracy dropped due to the lower accuracy of classes B and C. Similarly, on the right-side confusion matrices of Ph.D. model, classes A and C have higher accuracy than the B and D classes. The noises and outliers present in the collected dataset cause lower accuracy, especially in B and D classes, by making incorrect generalizations.

3.2.1. Prediction With most Important Feature

We can observe that for MS data, the highest accuracy of 69% has been achieved by the random forest method. For PhD dataset, a similar accuracy has been attained by random forest and decision tree algorithms. We tried to implement the k-fold cross validation technique [15] also, but unfortunately, the performance of the classification improved insignificantly. The desired result has not been obtained because of too many noisy data. We have done the analysis so far using the most important eight features, i.e., undergraduate CGPA and university, English test scores, research papers, and job experiences, described in the data analysis section. We are now finding out how many classifications are true and false according to the most significant features, undergraduate CGPA, and GRE. First of all, we have found out the average of undergraduate CGPA and GRE score. We find the average CGPA is 3.53, and the mean average GRE score is 300.1. Finally, we set four conditions to the undergraduate CGPA, and GRE score intuitively to find out the true and false class labeling. The conditions are:

1. If CGPA> Average CGPA and GRE > Average GRE, it is in class A.
2. Else if CGPA > Average CGPA and GRE < Average GRE, it is in class B.
3. Else if CGPA < Average CGPA and GRE> Average GRE, it is in class C.
4. Else if CGPA < Average CGPA and GRE < Average GRE, it is in class D.

Undergraduate CGPA is a higher important feature than the GRE score, as shown in Figure 6 and Figure 7. Out of 400 PhD data, we have found that only 140 data match these conditions, and for 300 MS data, it only matches 100 data, unfortunately. As the scope of admission and funding opportunities is uncertain, many Bangladeshi students with excellent academic profile apply to only individual universities, where the chances of admissions are almost definite for them. Our obtained dataset comprises many cases where the students have an outstanding academic background to get admission at class A universities, but actually, they only applied to lower-class universities. This section confirms this fact, and hence we can estimate the reasons behind not obtaining higher accuracy.

5. CONCLUSION (10 PT)

The undergraduate students from developing countries like Bangladesh invest a lot of money, time, and energy while applying for graduate studies. In this work, mathematical models have been developed to predict universities' admissions possibilities from the students' perspective. An individual student with his past academic records will be informed about which range of universities he should apply, and his chance of admission. We have used three algorithms, decision tree, random forest, and AdaBoost, for making a model for MS and PhD applicants separately. We found that the random forest classifier model for both the MS and PhD data outperformed all other models in terms of F1-score, accuracy, and AUC. This validates the reason behind other research papers used the random forest algorithm to anticipate graduate admissions. The decision tree algorithm offers similar prediction accuracy to the random forest for the PhD data but lower for the MS data. AdaBoost, the sequential ensemble learning method, performed poorly for both the MS and PhD students due to the noises and outliers in the dataset.

Many students in our acquired dataset applied and eventually got accepted in lower-ranking universities, although they have an outstanding profile, which leads to some noises and outliers. In the future, more data of the candidates of MS and PhD degrees can be collected and adopt a more proper way to divide them into different classes. We can increase the system's reliability by adding text features to the data, e.g., statements of purpose, research proposal, letters of recommendation, etc.

REFERENCES

[1] C. Romero and S. Ventura, Data Mining in e-Learning, WIT Press, 2006.
[2] "University Rankings," QS World University Rankings. [online]. Available: https://www.topuniversities.com/university-rankings/world-university-rankings/2021. [Accessed: 09-August-2020].
[3] A. K. Nandi, H. Ahmed, "Decision Trees and Random Forests," *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*, IEEE, 2019, pp.199-224.
[4] B. Schölkopf, J. Platt, and T. Hofmann, "AdaBoost is Consistent," *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, MIT Press, 2007, pp.105-112.
[5] J. M. Rosa, "Introduction to ΣΔ Modulators: Fundamentals, Basic Architecture and Performance Metrics," *Sigma-Delta Converters: Practical Design Guide*, IEEE, 2018, pp.1-27.
[6] N. T. N. Hien and P. Haddawy, "A decision support system for evaluating international student applications," *2007 37th Annual Frontiers in Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, Milwaukee*, WI, pp. F2A-1-F2A-6, 2007.
[7] A. Waters and R. Miikkulainen, "GRADE: Machine Learning Support for Graduate Admissions," AI Magazine, vol. 35, pp. 64-75, 2014.
[8] M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Chennai, India, pp. 1-5, 2019.
[9] I. Hmiedi, H. Najadat, Z. Halloush and I. Jalabneh, "Semi Supervised Prediction Model in Educational Data Mining," *2019 International Arab Conference on Information Technology (ACIT)*, Al Ain, United Arab Emirates, pp. 27-31, 2019.
[10] P. Janani, V. P. Hema, S. P. Monisha, "Prediction of MS Graduate Admissions using Decision Tree Algorithm," *International Journal of Science and Research (IJSR)*, vol. 9, pp. 492 – 495, March 2020.
[11] J. Huang, Y. F. Li, and M. Xie, "A Systematic Analysis of Data Preprocessing for Machine Learning- based Software Cost Estimation. Information and Software Technology", Elsevier, 2015, vol. 67, pp. 108-127.
[12] A. P. Cassidy and F. A. Deviney, "Calculating feature importance in data streams with concept drift using Online Random Forest," *2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, 2014, pp. 23-28.
[13] C. Jin, L. De-lin and M. Fen-xiang, "An improved ID3 decision tree algorithm," *2009 4th International Conference on Computer Science & Education*, Nanning, 2009, pp. 127-130.
[14] Y. Freund and R. E. Schapire, "A Short Introduction to Boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, September 1999, pp. 771-780.
[15] K. Pal and B. V. Patel, "Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2020, pp. 83-87.

BIOGRAPHIES OF AUTHORS (10 PT)

First author's Photo (3x4cm)	Xxxx (9 pt)
Second author's photo(3x4cm)	Xxxx (9 pt)
Thirth author's photo(3x4cm)	Xxxx (9 pt)