

Prediction Model on Student Performance based on Internal Assessment using Deep Learning

<https://doi.org/10.3991/ijet.v14i08.10001>

Sadiq Hussain ^(✉)

Dibrugarh University, Dibrugarh, India
sadiq@dibru.ac.in

Zahraa Fadhil Muhsin
Hup University, Baghdad, Iraq

Yass Khudheir Salal
South Ural State University, Chelyabinsk, Russia

Paraskevi Theodorou
National Center for Scientific Research “Demokritos”, Attica, Greece

Fikriye Kurtoğlu
Çukurova University, Adana, Turkey

G.C. Hazarika
Dibrugarh University, Dibrugarh, India

Abstract—Educational Data Mining and Deep learning play a crucial role in identifying academically weak students of an institute and help them by developing different recommendation systems to enhance their performance. These technologies direct the students for their future plan by discovering the precious hidden patterns from their history of information. Students from three colleges in Assam, India were considered in our research and their records were run on deep learning using the sequential neural model with the Adam optimization method. The study compared other classification methods such as the Artificial Immune Recognition System v2.0 and AdaBoost, to predict the results of the students. The highest classification accuracy achieved in this study was 95.34% produced by deep learning techniques. The Precision, Recall, F-Score, Accuracy, and Kappa Statistics Performance were calculated as a statistical decision to find the best classification methods. The dataset used in this study consisted of 10140 student records.

Keywords—Educational Data Mining (EDM), Deep Learning(DL), Classification, academically weak students, Artificial Immune Recognition System v2.0(AIRS2), Recurrent Neural Network (RNN), Adaptive Moment Estimation(Adam).

1 Introduction

Expecting the performance of the student activate is a substantial point in EDM which employs machine learning and data mining techniques to find out data from educational settings. Educational institutions required the result of the predicate process to improve their institute and improving student's academic achievement, this is done by analyzing a large volume of student data [1]. Figure 1 shows several factors that affect student performance [2].

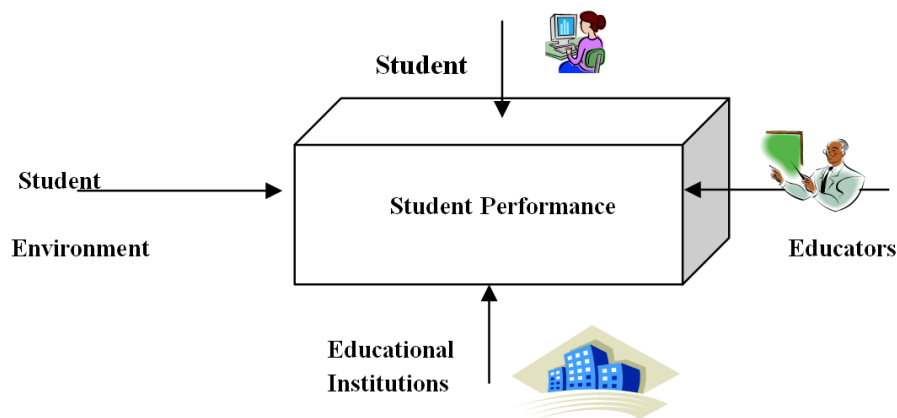


Fig. 1. Student Performance Factors

The main objects of learning analysis have been identified by [3]

- Monitoring and analysis of the learning process by the educational instructions. 2) Prediction or estimating learners' future knowledge or performance for supporting the required assistance.
- Guidance and advice mentoring
- Feedback by examining efficiency and effectiveness of learning, and providing meaningful feedback for teachers and learners.
- Modification the learning environments according to the results obtained.
- Recommendation by leaving the control to the learner.
- Self-reflection from the past learners' experiences or achievements. Figure 2 depicts the relationship of the Educational System with data mining and how students are benefited by such knowledge discovery.

However, there are so much potential and highly useful values hidden in the big data. Big data have many challenges, such as capturing, storage, analysis and visualization [4]. Educational organizations' database has been constantly increasing for many years and thus the volume becomes huge and this is an impediment on the analysis process. Researches required deep and enough knowledge to the big data for analyzing and studying the performance of the student and the institutes. Figure 2

depicts the relationship of the Educational System with data mining and how students are benefited by such knowledge discovery.

The rest of the paper is organized as section ii described the related works, section iii presented the methodology, section iv dealt with the dataset description and classifier evaluation techniques, section v focused on the experimental results, section vi presented the conclusion of the research work.

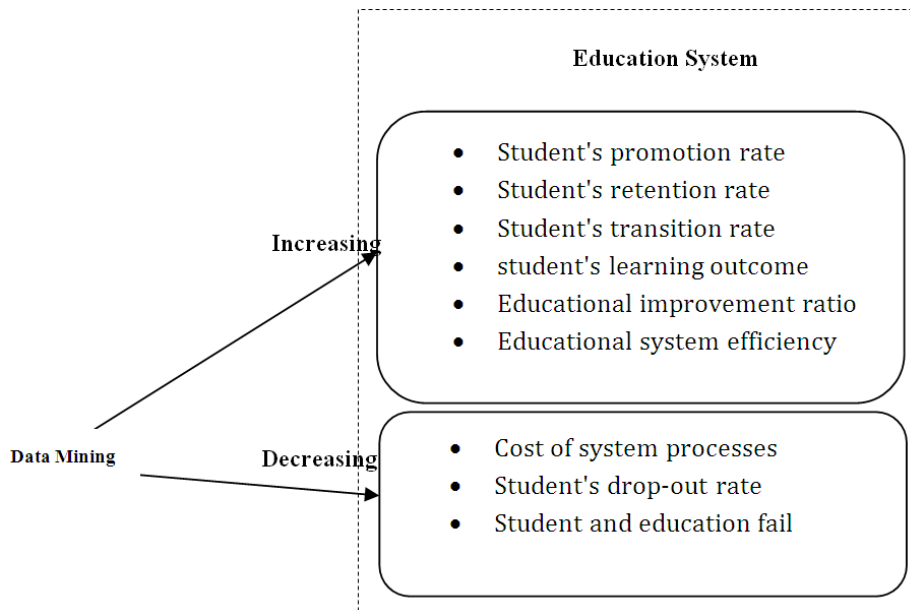


Fig. 2. Increasing and decreasing relation of Data Mining with Education System

2 Related Works

2.1 Educational data mining

The emergence of “Educational big data” (very large data sets compiled from students’ interactions with educational software and online learning) can improve learning processes in formal and informal education [5]. Educational Data Mining is an emerging discipline, which develops methods for exploring the unique types of data that comes from educational settings [6]. Data mining is the extraction of non-trivial, implicit and potentially useful patterns or knowledge and behavior from a huge amount of data. They can be interpreted and adjusted to check the efficiency of the e-learning systems such as Moodle [7]. Some of the most frequently used methods and techniques of data mining are Clustering, Classification and Sequential Analysis that can lead to the identification of useful patterns and to a reasonable prediction of future events in education [8]. Moreover, these methods aim to improve the understanding of the framework within which the learning process of students takes place [9].

2.2 Deep learning used by researchers

Gijbels et al [10], examined the relationship between students' approaches to learning and their measurable learning outcomes, in terms of problem-solving skills assessed by multiple-choice examination. The cognitive components of problem-solving were the understanding of principles, procedures etc.

Clare et al [11], described an attempt to involve the final year's social work students in an exchange of teaching, learning, and evaluation that promoted critical reflection by implementing a "deep learning" framework. The way this expertise was achieved remains, however, overlooked in an increasingly focused environment centered on the results, proofs of performance and teachers concerns about content and curriculum assessment.

2.3 Classification techniques

Sadiq Hussain et al [12], used data mining tools and classification techniques in order to improve the academic performance of students and to prevent dropping out. The classification methods that were used were the J48, PART, Random Forest and Bayes Network Classifiers and the data mining tool used was WEKA. According to Sadiq Hussain et al[12], "The internal assessment attribute in the continuous evaluation process makes the highest impact in the final semester results of the students in their dataset". Finally, the results showed that random forest outperformed the other classifiers based on accuracy and classifier errors.

López et al [13], proposed classification through a clustering approach to predict the final marks in a university course, based on the usage data of the Moodle Forum. The aim is to determine whether student participation in the course forum could be a good predictor for the final marks of the course and to examine whether the proposed classification through clustering approach can get the same precision as traditional classification algorithms.

Mashat et al [14], presented an efficient clustering technique by using the K-Means algorithm. The model used the K-Means algorithm and Davies-Bouldin index (DB) as an internal index for the clustering evaluation. The experimental results show that K-Means performance surpassed both Fuzzy C-Means and SOM algorithms.

Klašnja-Milićević et al [15], proposed a recommendation module of a programming tutoring system - Protus, which could be automatically adapted to learners' interests and levels of knowledge. This system recognized different patterns of learning styles and learners' habits by testing the learning styles of learners and mining their server logs. The results showed the suitability of using this recommendation model, in order to suggest online learning activities to learners based on their learning style, knowledge, and preferences.

Ayesha et al [16], applied a data mining technique to analyze student's learning behavior. They used the K-means clustering method to discover knowledge coming from the educational environment. They examined student's assessment factors, such as mid and final exams. The aim was to help the teachers to reduce the drop-out ratio to a significant level and improve student performance.

Alfianiand Wulandari in [17], focused on mapping students using the K-mean cluster algorithm to reveal the hidden pattern and classify students based on their demographic and average course attendance. Mapping the current student status was the requirement that needed to be made before student performance improved.

2.4 Clustering techniques

Sadiq Hussain et al [18], searched the association rules using data with data mining tools appropriate for the educational data as Orange, Weka and R Studio. They also used various clustering and classification methods in order to compare the suitable one for the dataset.

Campagni et al [19], proposed a data mining methodology based on clustering and sequential pattern analysis to study the behavior of university graduate students in their careers in order to identify strategies for improving the performance of students and the scheduling of exams by comparing student careers to the ideal one. The results showed that the more students follow the order given by the ideal career the more they achieve good performances in terms of final grade.

Bogarin et al [20], proposed to use data from Moodle's logs for each cluster of students separately in order to obtain more specific and accurate models of students' behavior. The results showed that the suitability of the specific model was greater than the general model obtained using all the data, and in some cases, the comprehensibility of the models could be improved.

Bovo et al [21], used data mining, machine learning and artificial intelligence methods to monitor students in e-learning training. Their aim was to create a system that would collect and process all data associated with a particular e-learning course. Furthermore, they examined whether there was an overall ideal number of clusters and whether the clusters showed mostly qualitative or quantitative differences. The clustering experiments were performed using actual data obtained from a Moodle platform.

Antonenko et al [22], applied a hierarchical clustering method (Ward's clustering) and a non-hierarchical clustering method (k-Means clustering) to analyze learning behavior characteristics while learners engaged in a problem-solving activity in an online learning environment. The use of cluster analysis helped to identify characteristics of more or less successful learning strategies and to understand the nature of cognitive and metacognitive processes that formed the basis of knowledge acquisition and provided guidance on how to perfect educational practices.

Tam et al [23], proposed an e-learning framework that can be applied to find an optimal learning sequence of the concepts or modules involved for each individual learner, mapping the desired learning sequences. The optimum learning sequence was generated after considering a student's past performance, learning profile and learning style, and conducting a semantic analysis of the course materials. This was followed by a heuristic-based concept clustering algorithm to return the optimal learning sequence. The proposed framework also included an examination of several experts' recommended learning sequences – however contradictory these might be.

3 Methodology

3.1 Data mining

First, the traditional data analysis is supposition operator, which means it begins from a question and explores the data to assure the conjecture; it uses analysis technologies that can interactive reconnoitering the data. It could start by a set of predefined reports that consider knowing interest directions. At the other hand, the question-driven process can be an important action to recognize the information when reasonable factors are participating, it can be hard to recognize more complex data that concern with different side [24].

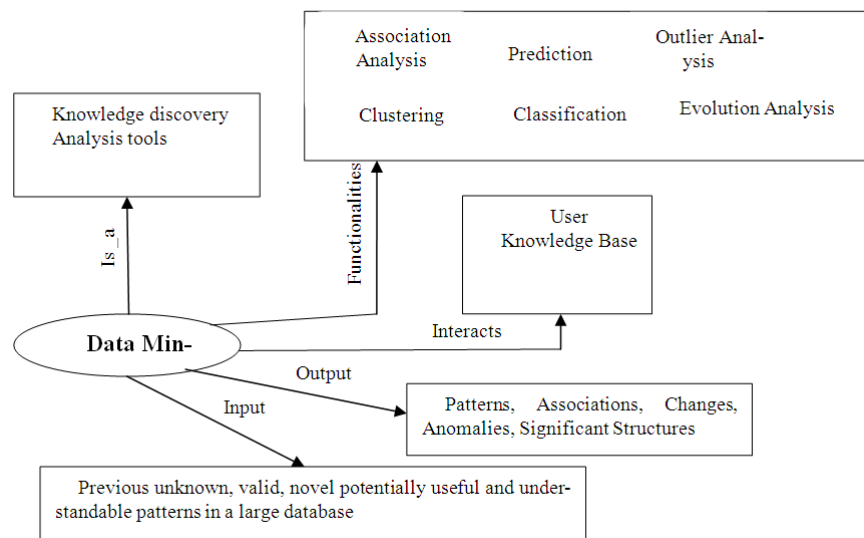


Fig. 3. Different Tasks Performed during the Data Mining Process

Instead of working with classic data analysis, data mining discovers significant patterns and destination by building analytic models. Data mining is not only an analytical tool that builds models but also a technological process in the real application. Currently, the concepts and techniques of data mining is a promising frontier in the database system and new database system applications. Data mining is considered a highly desirable tool being applied to many application areas as described in the figure 3.

3.2 Deep learning (DL) and recurrent neural network (RNN)

DL has been common and effective employed in a broad field of EDM by current researches. Linear separable areas and noisy complex input data obvious open the

area of working with DL. Currently, it is hopeless to move closer to understand how DL arrives at their conclusion. DL is so ambiguous (as a block box) for any clear performance to how they carried out their conclusion. Therefore DL is not effective when the conclusion required an explicit declaration for it. DL shows promising results in many applications in spite of its downsides. DL applied to education data mining in many studies. Beyond the Deep learning process an effective encoding of converting the information to future predictions [25].

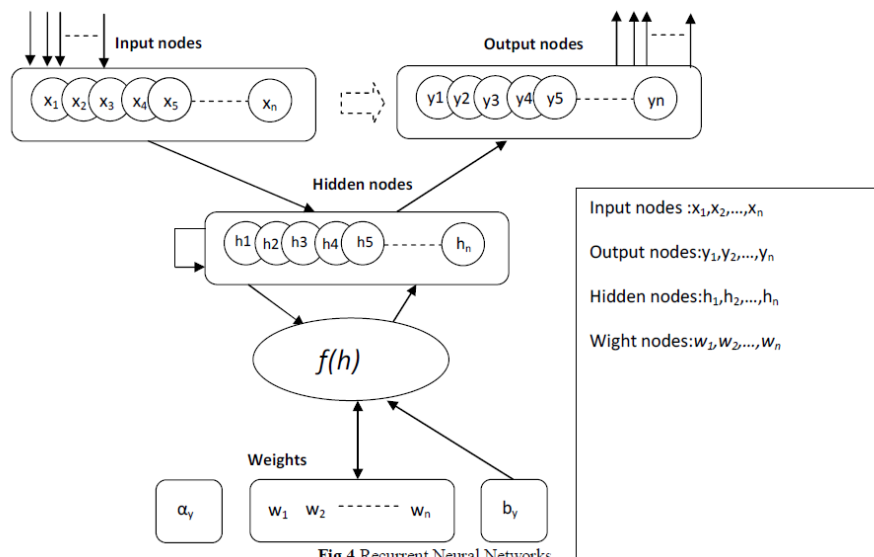


Fig.4.Recurrent Neural Networks
Fig. 4. Recurrent Neural Networks

Numerous learning algorithms have been added to the family of neural networks recently. Neural networks are techniques that imitating the human brain acts. Neural network characterized by parallel processing and the efficiency of deducing result from complex and ambiguous data unlike many other classification algorithms, identical to the human brain processing. Neural Networks can be discovered in confusing patterns that are complicated to be concluded by human brains or any learning algorithms. The cooperation of nodes in neural networks is varying among algorithms such as straightforward or backward (sequential or convolution). These nodes interconnected each other to exchange information, mimicking the way our neurons correlate through dendrites and axons for information interchange. Recurrent Neural Networks are the algorithm of neural networks that supervise the control of the information among the nodes. Recurrent Neural Networks hold out the learning process and complex transformations of data which require a long process time with sequential inputs. The figure shows the classic Recurrent Neural Networks which plot the sequence input vectors to the sequence output vectors which is done through computing of a sequence of hidden nodes [26].

3.3 Adam optimization methods

Adam (Adaptive Moment Estimation) is a successful and widespread algorithm in the domain of training deep or complex neural networks. Many pieces of research on the field of science and engineering choose the ADAM optimization because it encourages many current optimization methods and confers a faster convergence instead of standard stochastic gradient descent with moment [27]. This algorithm helps the researcher to override the disadvantage of the conventional. It meets the requirements of controlling the optimization, little memory requirement, straightforward implementation, efficient computation, the diagonal rescaling of the gradients is changeable, convenient with the huge data computation, suitable for non-stationary objectives, appropriate in troubles with unclear and scattered gradients, the hyperparameters commonly demand a few setting, interpretations prediction, reducing the disadvantage of the square hinge, avoiding the stuck in saddle points rather than local minima. Adam keeps an exponentially decaying average of past gradients [28].

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2)$$

m_t and v_t is an estimate of the moment of the gradients, where they initialized as vectors of 0's, β_1, β_2 closed to zero. They counteract these biases by computing bias-corrected moment estimates.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (4)$$

Then used the update rule

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \quad (5)$$

Default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, $\epsilon = 10^{-8}$.

3.4 AIRS2 (Artificial Immune Recognition System v2.0)

Many research inspired by the characteristics of the AIRS and several applications used this technology successfully. AIRS2 have received a lot of researcher's attention who were planning to establish models based on immune system and techniques in order to provide solutions for complicated problems [29-31].

The main idea of the AIRS2 is the application of the learning strategy to produce a core data point to shape the modified space for each distinct class. Therefore, these clarify points of data would be performed to design a molding of the education future. However, the good of this strategy remained on the core data points closely minutely chosen by the AIS system. While the AIS create memory cells, other methods only use these points for pattern predictions. AIRS2 used with supervised learning classifi-

cation problems and to substantiate tuberculosis diagnosis identification. AIRS2 is a class of adaptive or learning computer algorithm inspired by the function of the biological immune system, designed for applied to difficult problems such as classification[29-31].

The most characteristic of the AIRS2 provides a reduce memory cells pool (database) and overcome the limitations of assigning a class membership to each cell (the classified value). AIRS2 is a supervised learning algorithm. The mechanisms used in AIRS2 are resource competitions, Clonal selection, affinity maturation and memory cell formation [29-31].

3.5 Adaboost classification techniques

Ensemble classifiers is a technique that combines several classifiers to form a strong classifier, there have been several of research in ensemble modeling in artificial intelligence, statistics, technology, medicine in the last three decades; it is possible that this is because of the reliability and effectiveness in predictions compares with the others classifiers, ensemble classifiers called a meta-learning method that trains varied weak classifiers and appoints the knowledge of various and easy learners in making an assembly of classifiers. Ensemble classifiers have provided a power base for research in both the statistical communities and machine learning, this kind of classifier can improve classification accuracy. Multi-lingual systems like an ensemble of classifiers have attempted to ameliorate the accuracy of the general learning system. The AdaBoost classifier is an ensemble mechanism that trying to form a robust classifier from a set of weak classifiers (low accuracy), more recently it may be indicated to as discrete, actually because it is used for classification than regression and is best used to increase and improves the performance of decision trees to overcome the impediments of binary classification, and it is the best onset point for comprehension “boosting”. The AdaBoost algorithm needs to get feedback from the prior classification, and then hand out the weight of any classifier according to this feedback. So, the power of AdaBoost have been attributed to the algorithm’s ability to decrease the training error and speed up congregation after several iterations, which deals with many fields, including in Learning Analytics, can define the "Learning Analytics" as the gathering, measurement, analysis and preparing a report of data about the status of learners, for aims of understanding and improve learning [32-37].

4 Dataset Description Classifier Evaluation Techniques

4.1 Dataset description

The Dataset was collected from the three different colleges of Assam, India. They were Digboi College, Duliajan College, and Doomdooma College. There were 10140 records with 10 attributes. There were some missing values in the datasets. The missing values were not considered. Table 1 shows the description of the attributes in this research and descript there values. Figure 5,6 presented a pictorial view of the dataset.

Table 1. Attribute description with their values

| Attribute | Description | Values |
|-----------|--|--|
| Exam | The Three Year Degree Six-semester Exams | {'BA', 'BSC' } Two Examinations are taken into account, i.e. BA and BSc |
| Sub | Major/Honours Subjects Of BA and BSc Programmes | {'ENGM','PHYM', etc.} ENGM- Major/Honours in English PHYM- Major/Honours in Physics |
| IN_Sem1 | Internal Assessment Marks obtained in the BA/BSc 1st Semester Examination | Maximum marks 20 Marks obtained by the students in the range 1 to 20. Mean: 15.66257 Standard Deviation: 2.593816 |
| IN_Sem2 | Internal Assessment Marks obtained in the BA/BSc 2nd Semester Examination | Maximum marks 20 Marks obtained by the students in the range 1 to 20. Mean: 15.86548 Standard Deviation: 2.740797 |
| IN_Sem3 | Internal Assessment Marks obtained in the BA/BSc 3rd Semester Examination | Maximum marks 40 Marks obtained by the students in the range 1 to 40. Mean: 31.95765 Standard Deviation: 5.101312 |
| IN_Sem4 | Internal Assessment Marks obtained in the BA/BSc 4th Semester Examination | Maximum marks 40 Marks obtained by the students in the range 1 to 40. Mean: 30.80859 Standard Deviation: 5.43647 |
| IN_Sem5 | Internal Assessment Marks obtained in the BA/BSc 5th Semester Examination | Maximum marks 80 Marks obtained by the students in the range 1 to 80. Mean: 64.71536 Standard Deviation: 10.18944 |
| IN_Sem6 | Internal Assessment Marks obtained in the BA/BSc 6th Semester Examination | Maximum marks 80 Marks obtained by the students in the range 1 to 80. Mean: 64.79921 Standard Deviation: 10.3252 |
| InPc | The overall percentage secured by the candidate in all the six semesters in the internal assessments | Mean: 80.44676 Standard Deviation: 11.01706 |
| Result | The overall result of the candidate based the all the six semesters theory and internal assessment | {'Pass', 'Fail'} If a student secures 40% or above, he is termed as 'Pass' Else 'Fail' |

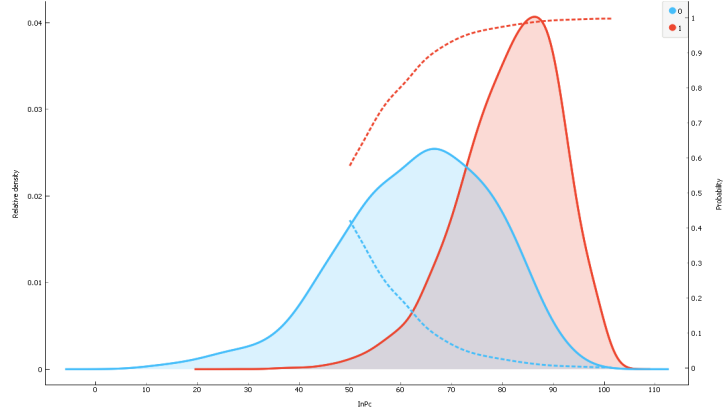


Fig. 5. Distributions of internal assessment percentage of the students with respect to results

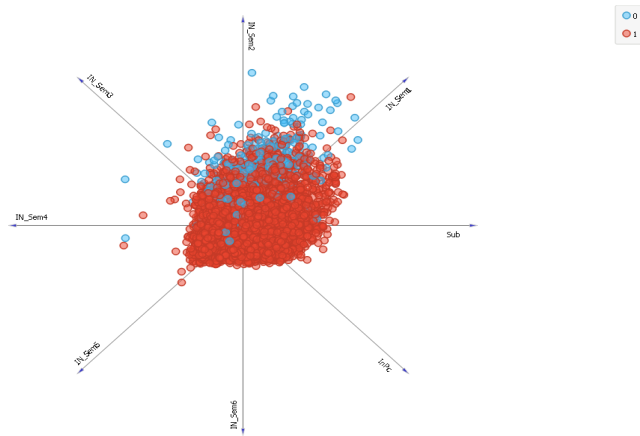


Fig. 6. Linear projection of all the features with respect to results

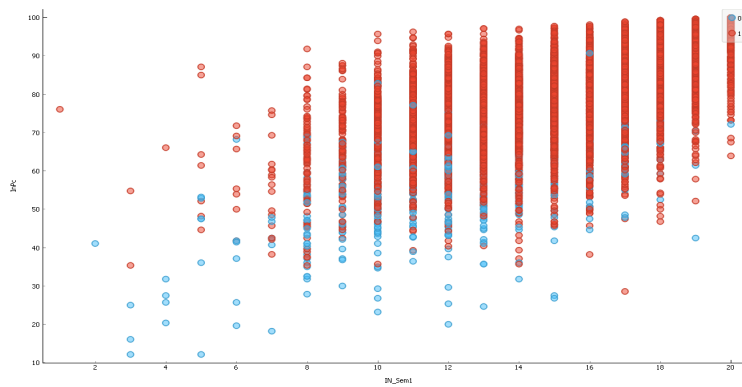


Fig. 7. Scatter Plot of 1st Semester Internal Assessment (IA) marks and Overall IA Percentage in six semesters with respect to results

4.2 Classifier evaluation techniques

Precision (ratio of positive documents among retrieved ones), Recall (ratio of retrieved positive documents among all positive documents) and Accuracy: The performance in the task was measured by calculating the break-even point which is a hypothetical point where precision (ratio of positive documents among retrieved ones) and recall (ratio of retrieved positive documents among all positive documents) meet when varying the threshold [38].

In the information retrieval community, the performance of an ad hoc retrieval system is typically evaluated using relevance-based recall and precision at a certain cut-off of the ranked result. Taking a 20-document cut-off as an example:

$$\text{precision} = \frac{\text{the number of relevant documents among the top 20}}{20}$$

$$\text{recall} = \frac{\text{the number of relevant documents in the top 20}}{\text{all relevant documents in the corpus}}$$

What is a good cut off number is unknown? In order to compare different algorithms without a specific cut off, the mean of the precision scores after each relevant document is retrieved, which is called Mean Average Precision, is often used [38].

In classification problems, the primary source of performance measurements is a coincidence matrix (a.k.a. classification matrix or a contingency table). Figure 7 shows a coincidence matrix for a two-class classification problem. The equations of most commonly used metrics that can be calculated from the coincidence matrix are also given below [39].

| | | True Class | |
|-----------------|----------|---------------------------|---------------------------|
| | | Positive | Negative |
| Predicted Class | Positive | True Positive Count (TP) | False Positive Count (FP) |
| | Negative | False Negative Count (FN) | True Negative Count (TN) |

Fig. 8. A simple coincidence matrix

The numbers along the diagonal from upper-left to lower-right represent the correct decisions made, and the numbers outside this diagonal represent the errors. The true positive rate (also called hit rate or recall) of a classifier is estimated by dividing the correctly classified positives (the true positive count) by the total positive count. The false positive rate (also called false alarm rate) of the classifier is estimated by dividing the incorrectly classified negatives (the false negative count) by the total negatives. The overall accuracy of a classifier is estimated by dividing the total correctly classified positives and negatives by the total number of samples. Other performance measures, such as recall (a.k.a. sensitivity), specificity and F-measure are also used for calculating other aggregated performance measures (e.g., area under the ROC curves) [39].

$$\begin{aligned} \text{accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\ \text{precision} &= \frac{TP}{TP+FP} \\ \text{recall} &= \frac{TP}{TP+FN} \\ \text{F - measure} &= \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} \end{aligned}$$

F-measure: F-measure (or F-score or F_1 -score) has been introduced to balance between sensitivity (recall) and specificity (precision). It is defined as the harmonic mean of the two scores, multiplied by 2 to obtain a score of 1 when both sensitivity and specificity equal 1 [39]:

$$F = 2 \cdot \frac{1}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Care must be taken when selecting an accuracy measure in presence of unbalanced classes [40].

Cohen's Kappa statistic: Cohen's Kappa statistic is a very useful, but under-utilized, metric. Sometimes in machine learning, we are faced with a multi-class classification problem [41]. In those cases, measures such as the accuracy or precision/recall do not provide the complete picture of the performance of our classifier.

In some other cases, we might face a problem with imbalanced classes. E.g. we have two classes, say A and B, and A shows up on 5 of the time. Accuracy can be misleading, so we go for measures such as precision and recall. There are ways to combine the two, such as the F-measure, but the F-measure does not have a very good intuitive explanation, other than it being the harmonic mean of precision and recall [41].

Cohen's kappa statistic is a very good measure that can handle very well both multi-class and imbalanced class problems.

Cohen's kappa is defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e}$$

where p_o is the observed agreement, and p_e is the expected agreement. It basically tells you how much better your classifier is performing over the performance of a classifier that simply guesses at random according to the frequency of each class.

Cohen's kappa is always less than or equal to 1. Values of 0 or less indicate that the classifier is useless. There is no standardized way to interpret its values. Landis and Koch [42] provide a way to characterize values. According to their scheme a value <0 is indicating no agreement, 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect agreement [42].

5 Experimental Results

Classification using deep learning was used to interrogate the dataset. The sequential neural model and the Adam optimization method were applied. The loss function used was `binary_crossentropy`. The first two layers used the rectifier activation function. The activation function used in the output layer was sigmoid. The first layer consisted of 12 neurons and 9 input variables. The hidden layer had 8 neurons and the output layer consisted of 1 neuron for prediction of the class (whether pass or fail). To implement it, Keras with TensorFlow was applied. For the deep learning model, test data were taken as 33%. The other classification methods were also applied. Table 2 gives a comparison of all the methods. For artificial immune recognition system, v2.0 and Adaboost classifiers 10 fold cross-validation methods were applied. For classification based on deep learning, numbers of Epochs used were 150 and the batch size was 10. Total and trainable parameters were found to be 247. The deep learning classification achieved higher accuracy than its counterparts.

Table 2. Comparison of different classification methods

| Classifier | Precision | Recall | F-Score | Accuracy | Kappa Statistics |
|---|-----------|--------|---------|----------|------------------|
| Deep Learning | 0.96 | 0.99 | 0.98 | 95.34 | 0.26 |
| Artificial Immune Recognition System v2.0 | 0.926 | 0.932 | 0.929 | 93.18 | 0.21 |
| Adaboost | 0.934 | 0.946 | 0.939 | 94.57 | 0.28 |

The followings depicted part of the experimental results of the classification using deep learning.

```
Epoch 140/150
10140/10140 [=====] - 1s 97us/step -
loss: 0.1445 - acc: 0.9552
Epoch 141/150
10140/10140 [=====] - 1s 99us/step -
loss: 0.1439 - acc: 0.9560
Epoch 142/150
10140/10140 [=====] - 1s 97us/step -
loss: 0.1437 - acc: 0.9550
Epoch 143/150
10140/10140 [=====] - 1s 97us/step -
```

```

loss: 0.1442 - acc: 0.9565
Epoch 144/150
10140/10140 [=====] - 1s 99us/step -
loss: 0.1447 - acc: 0.9554
Epoch 145/150
10140/10140 [=====] - 1s 97us/step -
loss: 0.1456 - acc: 0.9554
Epoch 146/150
10140/10140 [=====] - 1s 99us/step -
loss: 0.1458 - acc: 0.9553
Epoch 147/150
10140/10140 [=====] - 1s 97us/step -
loss: 0.1457 - acc: 0.9549
Epoch 148/150
10140/10140 [=====] - 1s 99us/step -
loss: 0.1445 - acc: 0.9552
Epoch 149/150
10140/10140 [=====] - 1s 97us/step -
loss: 0.1441 - acc: 0.9549
Epoch 150/150
10140/10140 [=====] - 1s 107us/step -
loss: 0.1440 - acc: 0.9561
4995/4995 [=====] - 0s 53us/step
accuracy: 95.34%
    
```

Table 3. Model summary

| Layer (type) | Output Shape | Param # |
|------------------|--------------|---------|
| dense_28 (Dense) | (None, 12) | 120 |
| dense_29 (Dense) | (None, 9) | 117 |
| dense_30 (Dense) | (None, 1) | 10 |

6 Conclusion

Research in the higher education system leads to improve the educational organization. The prediction of the student's success helps the education organization to supply the student with additional assessment; this process also enhances the development of the education system in their organization. From the experimental results, by going through the internal assessment of a candidate the result of the candidate might be predicted. So, such candidates with poor internal assessment marks, the tutors may devote some extra time to improve the final examinations marks. The students whose performance was poor internal assessment and they might be classified as at-risk students. As the predictive model might be used to warn such students and parents, it might be applied to improving the marks of such candidates. The instructors might intervene in real time by looking at the internal assessment marks secured by the stu-

dents. So, the internal assessment may be continued as an integral part of a particular course. The application of Machine Learning Techniques in predicting students' performance proved to be helpful for identifying poor performers and it can enable tutors to take remedial measures at an earlier stage, even from the very beginning of an academic year using only students' internal assessment data of previous semesters, in order to provide additional help to the groups at risk.

7 Acknowledgement

The authors acknowledged the Principals of Digboi College, Doomdooma College and Duliajan College of Assam, India for allowing us for using their data for the research purpose.

8 References

- [1] Agrawal R.S, Pandya M.H, 2015, "Survey of papers for Data Mining with Neural Networks to Predict the Student's Academic Achievements", International Journal of Computer Science Trends and Technology (IJCT), V3, I5, Sep-Oct 2015.
- [2] M. R. Beikzadeh and N. Delavari, "A New Analysis Model for Data Mining Processes in Higher Educational Systems," On the proceedings of the 6th Information Technology Based Higher Education and Training 7-9 July 2005.
- [3] Steiner, C., Kickmeier-Rust, M., & Albert, D. (2014). Learning Analytics and Educational Data Mining: An Overview of Recent Techniques. Learning analytics for and in serious games (pp. 6-15).
- [4] C.L. Philip Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques, and technologies: A survey on Big Data, Inform. Sci. (2014). <https://doi.org/10.1016/j.ins.2014.01.015>
- [5] Siemens, G., & Baker, R.S. (2012). Learning analytics and educational data mining: towards communication and collaboration.
- [6] Aher, S.B. (2011). Data Mining in Educational System using WEKA.
- [7] Sunita B Aher and Lobo L.M.R.J. Mining Association Rule in Classified Data for Course Recommender System in E-Learning. International Journal of Computer Applications 39(7):1-7.
- [8] Felix, I.M., Ambrosio, A.P., Neves, P.S., Siqueira, J., & Brancher, J.D. (2017). Moodle Predicta: A Data Mining Tool for Student Follow Up. CSEDU.
- [9] International Educational Data Mining Society www.educationaldatamining.org.
- [10] Gijbels, D., Van de Watering, G., Dochy, F., & Van den Bossche, P. (2005). The relationship between students' approaches to learning and the assessment of learning outcomes. Euro-pean Journal of Psychology of Education, 20(4), 327-341. <https://doi.org/10.1007/BF03173560>
- [11] Clare, B. (2007). Promoting Deep Learning: A Teaching, Learning and Assessment Endeavour. Social Work Education, 26(5), 433-446. <https://doi.org/10.1080/02615470601118571>
- [12] Hussain, S., Dahan N. A., Ba-Alwib F. M., Najoua R (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. Indonesian Journal of Electri-

- cal Engineering and Computer Science, Vol. 9, No.2, <https://doi.org/10.11591/ijeecs.v9.i2.pp447-459>
- [13] López, M. et al., 2012. Classification via clustering for predicting final marks based on student participation in forums. Proceedings of the 5th International Conference on Educational Data Mining, pp.4–7.
- [14] Mashat, A.F. et al., 2012. Efficient Clustering Technique for University Admission Data. International Journal of Computer Applications, 45(23), pp.39–42.
- [15] Klačnja-Milićević, A. et al., 2011. E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, Vol. 56, No. 3, pp.885–899. <https://doi.org/10.1016/j.compedu.2010.11.001>
- [16] Ayesha, S. et al., 2010. Data Mining Model for Higher Education System. European Journal of Scientific Research, Vol. 43, No.1, pp.24 – 29.
- [17] PermataAlfiani, A. & AyuWulandari, F., 2015. Mapping Student's Performance Based on the Data Mining Approach (A Case Study). Agriculture and Agricultural Science Procedia, 3, pp.173–177. <https://doi.org/10.1016/j.aaspro.2015.01.034>
- [18] Hussain, S., Atallah, R., Kamsin, A., & Hazarika, J. (2018). Classification, Clustering and Association Rule Mining in Educational Datasets Using Data Mining Tools: A Case Study. Advances in Intelligent Systems and Computing, 196-211. doi:10.1007/978-3-319-91192-2_21. https://doi.org/10.1007/978-3-319-91192-2_21
- [19] Campagni, R. et al., 2015. Data mining models for student careers. Expert Systems with Applications, Vol. 42, No. 13, pp.5508–5521. <https://doi.org/10.1016/j.eswa.2015.02.052>
- [20] Bogarín, A. et al., 2014. Clustering for improving educational process mining. Proceedings of the Fourth International Conference on Learning Analytics and Knowledge - LAK '14, pp.11–15. <https://doi.org/10.1145/2567574.2567604>
- [21] Bovo, A. et al., 2013. Clustering moodles data as a tool for profiling students. In 2013 Second International Conference on E-Learning and E-Technologies in Education (ICEEE). pp. 121–126. <https://doi.org/10.1109/ICeLeTE.2013.6644359>
- [22] Antonenko, P.D., Toy, S. & Niederhauser, D.S., 2012. Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, Vol. 60, No. 3, pp.383–398. <https://doi.org/10.1007/s11423-012-9235-8>
- [23] Tam, V., Lam, E.Y. & Fung, S.T., 2012. Toward a complete e-learning system framework for semantic analysis, concept clustering and learning path optimization. Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies, ICALT 2012, pp.592–596. <https://doi.org/10.1109/ICALT.2012.66>
- [24] E. Gaudioso, L. Talavera, "Data mining to support tutoring in virtual learning communities: experiences and challenges", CHAPTER 12, 2006.
- [25] Bendangnuksung, Prabu P, Students' Performance Prediction Using Deep Neural Network, International Journal of Applied Engineering Research ISSN 0973-4562 V.13, N.2, 2018, p. 1171-1176.
- [26] Peter Wlodarczak, Jeffrey Soar, Mustafa Ally, "Multimedia Data Mining using Deep Learning", IEEE 2015, p190-196.
- [27] Diederik P. Kingma, Jimmy Lei Ba, ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION, In the Conference Proceedings of ICLR 2015.
- [28] Ruder Sebastian, An overview of gradient descent optimization algorithms, arXiv:1609.04747v2 [cs.LG] 15 Jun 2017.
- [29] Peng, Y., & Lu, B. (2015). Hybrid learning clonal selection algorithm. Inf. Sci., 296, 128-146. <https://doi.org/10.1016/j.ins.2014.10.056>

- [30] Shamsheerband S, Hessam S, Javidnia H, Amiribesheli M, Vahdat S, Petković D, Gani A, Kiah MLM. Tuberculosis Disease Diagnosis Using Artificial Immune Recognition System. *Int J Med Sci* 2014; 11(5):508-514. <https://doi.org/10.7150/ijms.8249>
- [31] Saidi, M., Chikh, A., &Settouti, N. (2011). Automatic Identification of Diabetes Diseases using an Artificial Immune Recognition System2 (AIRS2) with a Fuzzy K-Nearest Neighbor. *CIIA*.
- [32] ADEGOKE, Vincent F., et al. Prediction of breast cancer survivability using ensemble algorithms. In: *Smart Systems and Technologies (SST), 2017 International Conference on*. IEEE, 2017.p. 223-231.
- [33] V. Wheway, Using Boosting to Detect Noisy Data, *PRICAIWorkshops,2000*, pp.123-132.
- [34] DIETTERICH, Thomas G. Ensemble methods in machine learning. In: *International workshop on multiple classifier systems*. Springer, Berlin, Heidelberg, 2000.p. 1-15.
- [35] HAN, Meimei, et al. Application of Ensemble Algorithm in Students' Performance Prediction. In: *Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress on*. IEEE, 2017.p. 735-740.
- [36] I. Mukherjee, C. Rudin, and R. Schapire, "The rate of convergence of AdaBoost," *Journal of Machine Learning Research*, vol. 14, pp.2315-2347, 2013.
- [37] SIEMENS G, LONG P. Penetrating the fog: Analytics in learning and education. *Educuse Review*, 2011, 46(5): 30-32.
- [38] Srivastava, A., Sahami, M. (2009). *Text mining: classification, clustering, and applications*. Chapman and Hall, Boca Raton. <https://doi.org/10.1201/9781420059458>
- [39] Olson, D.L., Delen, D. (2008). *Advanced Data Mining Techniques*. Springer-Verlag Berlin.
- [40] Galdi, P., Tagliaferri, R. (2019). Data Mining: Accuracy and Error Measures for Classification and Prediction. In book: *Reference Module in Life Sciences, Encyclopedia of Bioinformatics and Computational Biology, Volume 1, Pages 431-436*.
- [41] <http://thedata scientist.com/performance-measures-cohens-kappa-statistic/>
- [42] Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33(1):159-174. <https://doi.org/10.2307/2529310>

8 Authors

Dr. Sadiq Hussain of Dibrugarh University, Assam has been a greater researcher and has published around 18 papers which have got Google citations. The first paper was published in 2014 on Educational Data Mining model using rattle.

Zahraa Fadhil Muhsin an eminent Nuclear Scientist emerging from Iraq completed his Nuclear Physics in 2007 from Baghdad University and persuaded his Post Doctorial Research in Malaysia University in 2012. Now he holds the chair of Dean in the University.

Yass Khudheir Salal is from the Department of System Programming, South. Ural State, Chelyabinsk in Russia.

Paraskevi Theodorou is from National Center for Scientific Research "Demokritos", AgiaParaskevi, Attica, Greece.

Dr. Fikriye Kurtoğlu an young dynamic scientist from Turkey has immense skills in Data Analysis, Statistical Analysis, Statistical Modeling, Applied Statistics, Statistics, R Programming, R Statistical Package, Linear Regression, Regression Analysis,

Maximum Likelihood, Generalized Linear Models. She has 6 research items amongst which 5 obtained citations

Dr. Gopal. C. Hazarika a mathematician from Assam's Dibrugarh University has published numerous titles and received 217 citations, in 2017 alone around 46 papers / titles got published.

Article submitted 2018-12-13. Resubmitted 2019-02-13. Final acceptance 2019-02-15. Final version published as submitted by the authors.