Improving Students' Performance Prediction using LSTM and Neural Network

Hussam Abduljabar Salim Ahmed[1], Razieh Asgarnezhad[2]

1- Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
Email: hussamj329@gmail.com

2- Department of Computer Engineering, Aghigh Institute of Higher Education Shahinshahr, 8314678755, Isfahan, Iran.
Email: razyehan@gmail.com (Corresponding author)

Received: 27 January 2023 Revised: 23 March 2023 Accepted: 3 April 2023

ABSTRACT:
Educational data mining utilizes information from academic fields to develop renewed techniques and spot unusual patterns to gauge students' academic achievement. Evaluating student learning is a complicated issue. Data mining in this field enables to predict students' performance to recommend performance in universities. Therefore, the current authors have recently seen the rapid growth of data mining and knowledge extraction as tools used by academic institutions to optimize student learning processes. Here, a method based on a certain kind of artificial neural network called Long Short Term Memory recurrent neural network for prediction will operate. The proposed approach tries to use the educational characteristics of different people to predict the best educational process future educational. It career for students and thereby take steps to improve the effectiveness of the educational system. For comparison, one of the newest algorithms presented in this field was implemented using the proposed technique. The evaluations' findings were performed in the form of two scenarios with different data sizes and different amounts of test and training data. For the evaluation, the dataset taken from an online educational system was used. The evaluation results are presented in the form of four well-known criteria precision, recall, accuracy, and F1, which demonstrate the superiority of the proposed method.

KEYWORDS: Educational data mining, LSTM, Artificial neural network, Deep learning

1. INTRODUCTION

Data mining uses many techniques and algorithms to discover patterns from recorded data. Researchers investigate these algorithms in various fields and suggest potential directions for further research.

Data mining in this field enables us to predict students' performance to recommend performance in universities. Therefore, there has been rapid growth in the use of data mining and knowledge extraction as tools used by academic institutions to optimize student learning processes [1,2]. Educational data mining is the name of this emerging discipline (EDM). To find new ways for education data, it considers the value of learning systems, education data mining produced [3]. In educational settings, predicting student academic success is crucial.

An incorrect admissions decision might guide to a student from being admitted to the institution. As a result, the quality of the approved applicants influences the institutions' degree of quality.

Predictive results can also be used by faculty to determine the most appropriate instructional measures for each group of students and equip more aid tailored to their needs. Accurate forecasting of students' progress is one of the ways to increase the level of quality and supply better educational services. Hence, it is important to provide an accurate forecasting tool for education institutions [4].

Although student performance is essential to the learning process, it is even a complex phenomenon that depends on many elements, including the teaching environment and individual study habits. Different factors have been operated in diverse study of indicators to produce models that can forecast academic success [5].

The history and basis of neural networks (NNs) obtained from a thorough and transparent examination of the brain on conventional computers [6]. Due to the nature of parallel computing, the NN is faster than a normal computer.

NN architecture is usually divided into two categories: Feedforward NN and NN with feedback. Three layers like input layer, one or more hidden layers and one output layer help in modeling a structure for the

Paper type: Research paper
DOI: 10.30486/mjtd.2023.1995783.1050
How to cite this paper: H. Abduljabar Salim Ahmed and R Asgarnezhad, "**Improving Students' Performance Prediction using LSTM and Neural Network**", *Majlesi Journal of Telecommunication Devices*, Vol. 12, No.3, pp.121-127, 2023.

NN [7]. The complexity of the system and the study specifies the number of hidden layers and neurons. Artificial neural networks (ANNs) have been used to indicate student success in many articles.

ANNs are basic computing methods that may easily mimic the functions of an animal's brain to tackle complicated issues. [6].

Deep learning uses the Long Short Term Memory (LSTM) architecture of artificial recurrent neural networks (ARNNs) [21]. LSTMs have feedback connections as opposed to typical feedforward NNs.

LSTM-based recurrent NN approach to predict student behavior are going to be to achieve this study. The proposed approach attempts to use the educational characteristics of different people to predict the best educational process as well as future educational and career for students and thereby take steps to enhance the performance of the educational system. In some existing methods, an approach including combining single link classification and LSTM neural network has not been used so far. In addition, the use of LSTM neural network for each category increase the quality of predictions is one of the innovations of the suggested method.

Due to the possibility of unpredictable time delays between significant time-series events, the LSTM network is effective for classifying, processing, and forecasting time series data. The vanishing gradient issue might arise when train conventional RNNs addressed by LSTMs. In many cases, LSTMs have a benefit over RNNs, hidden Markov models, and other sequencing learning techniques due to their relative insensitivity to gap length. [22].

The rest of this article is of concern: A review of the existing works shows in Sect. 2. Then, the proposed method is presented in Sect. 3. The obtained results presented in Sect. 5. Finally, conclusions and future suggestions are presented in Sect. 5.

2. RELATED WORK

In recent years, educational data mining has drawn a lot of interest. Some data mining approaches have been developed to remove hidden knowledge from educational data. The knowledge that is retrieved allows institutions to enhance both the teaching and learning processes. These all result in improved student performance and general learning outcomes [8].

Long-term dependencies may be learned utilizing specific RNNs called LSTM networks. These networks were first presented by Hochreiter and Schmidhuber in 1997 [9].

In actuality, the long-term reliance issue was the motivation behind the construction of LSTM networks. It is significant to mention that LSTM networks naturally and automatically memorize information for comprehensive periods and that one of their structural features allow them to effectively learn information that is extremely far from them.

In standard RNNs, these iterative modules have a simple structure, for example containing only one layer of hyperbolic tangent. Repetitive modules in standard RNNs have only one layer. LSTM networks have a similar. The iterative module has a different structure than a sequence or chain. They feature four layers instead of just one, and each layer interacts with the others in accordance with a unique structure.

In online learning environments, indicating student success is a crucial challenge. Several data mining techniques have been applied to design a predictive model. Classification is the most usually accustomed method for forecasting student achievement.

In paper [10], the demographic characteristics of students and CGPA for the first semester of postgraduate studies are used as a predictor variable for students' academic performance. Three prediction models have been developed, decision tree (DT), linear regression (LR), and ANN. The outcome of this study indicates that all three models produce accuracy levels of at least 80%. It reveals that the ANN outperforms the other two models as well. Correlation coefficient analysis is performed to choose the relationship between independent variables.

Romero et al. in [11] proposed a method based on information from their utilization in the model system. They utilized the DT method to forecast the final grades of the students. One of the most popular learning content management systems is Moodle (LCMS). The author divided students into two categories—passing and failing—using real data from seven Moodle courses at the University of Cordoba. This study's goal is to divide students into groups with similar final scores depending on the tasks they do in an online course.

Arsad et al. in [12] used the ANN model to estimate how well engineering bachelor students will succeed academically. The cumulative grade point average (CGPA) and deemed the study's output, comprised GP grading of the core disciplines, students regardless of their demographic background judged to be input. Engineering students are trained to achieve certain objectives using the GP neural network (NN). This research revealed how fundamental problems significantly affect the final CGPA after graduation.

The authors in [13] CGPA was predicted using Bayesian networks based on application data at the time of admission. Today's educational institutions require a system for determining whether candidates are qualified to get degrees from various universities. This study presents a novel method that integrates a case-based element with a prediction model. The former student, who is considered more like an applicant, is recovered by the case-based component. It is challenging to describe the items' (applicants') similarity in a way that is in line with the forecasting model. Any institution that possesses a reliable database of application and student data can utilize this strategy.

To assess and recommend students' academic achievement, they solve linear programming tasks of the appropriate difficulty level. Abu Nasser employed ANNs and expert systems to comprehend about the learning model in the linear programming intelligent teaching system. [14]. To predict academic performance, the Feed Backpropagation algorithm was trained with a set of learners' data. The accuracy of the learners' performance predictions was very high, and so he says the A excellent forecast may be made via an ANN.

Kanakana and Ulanrujajo [15] examined data from ANNs and LR models to predict students' achievement in higher education. Information from Tshwane University was utilized. The majority of the time, input variables were students with the twelfth-grade mean point total. In comparison to the LR values, the findings indicate a superior agreement between the predictions of the ANN model and the practical values.

Quindt et al. in [16] Using NN analysis, general academic success in the undergraduate academic year was anticipated based on student motivation, learning styles, working memory capacity, and attentiveness. 128 college students took part in this study. The findings revealed that working memory capacity and attention are both reliable indicators of academic success, particularly for the group's top and bottom performers. The desire and method of study of a group of students whose performance was on average 60% was a good predictor.

Mukta and Usha [17] utilized classic statistical methods and NNs to perform analysis to forecast the scientific performance of business school graduates, and the results were compared to gauge the effectiveness of these methods. Further, the fundamental components of the standard business school curriculum were defined, and their connections to the various aspects of the admissions process were laid out.

Stamos and Andreas [18] utilized a synthetic nerve to forecast the graduation rates of the pupils. The network was developed as a three-layer perceptron and was introduced through backpropagation techniques. Many experiments were conducted as training and testing exercises. A sample of 1,407 student profiles was utilized in these trials. The Woubonsee College pupils, for instance, were divided into two groups. 307 of the remaining 1100 profiles were used for testing, while the initial set of 1100 profiles was operated for training. For the practice and test sets, the average predictions were 77% and 68%, respectively.

In the paper [19], Using data mining techniques and additional data features known as student behavioral characteristics, a novel student performance prediction model is presented. The interaction of the learner with the e-learning management system is connected to these qualities. A group of categorizers, an ANN, a Bayesian network, and a DT are used to assess the performance of the student prediction model. Besides, group approaches have been applied to improve these categories' performance. Bagging, Boosting, and RF, which are popular group approaches employed in the literature, are employed in this research. The findings demonstrate a significant association between learner behavior and academic success. The precision of the suggested model has been improved up to a maximum of 22.1% by using behavioral features compared has achieved its highest level of accuracy by omitting these features from the findings of 25.8% using group methods. By testing the model using new students, the accuracy obtained is more than 80%. This result confirms the reliability of the proposed model.

3. THE PROPOSED METHOD

3.1. Risk Factor for Heart Disease

The proposed method, as remarked, consists of 2 different steps. Each of these steps is described below. Each of these steps is described in Figure 1.

In current paper, an LSTM-based approach is used to indicate student behavior. The suggested method in this paper. There are two processes involved in predicting kids' academic achievement of education and testing.

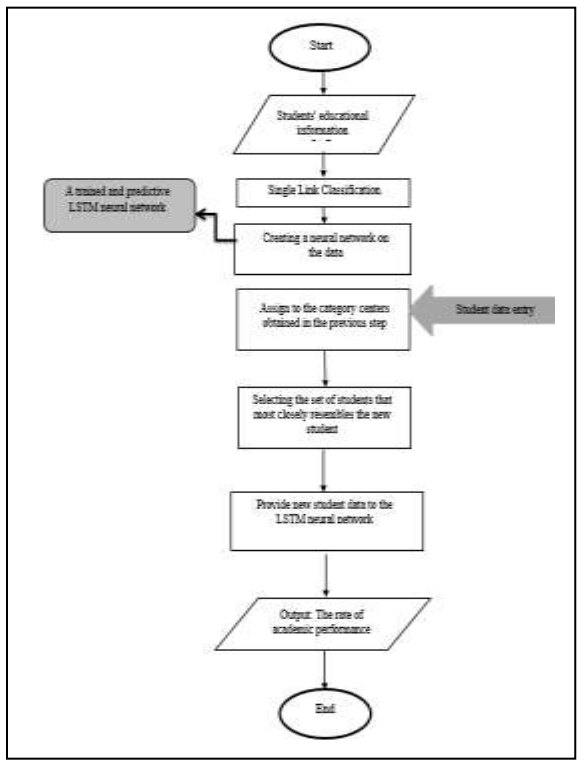

Fig. 1. Flowchart of the method.

3.2. Training Step

The training step consists of two parts: The classification part and the forecasting part.

Part One: Single Link Classification. In the training step, the acquired data contains the student's educational information, is first separated into different categories using the Single Link classification method. This classification, which is the basis of the suggested method and one of the innovations of this article, aids determine students' performance more accurately. Classifying students in this step will help similar students fall into the same categories and be distinguished by characteristics. As shown in the Single Link algorithm, the data can be divided into different categories according to the space. They are located, so that the data within a category is most similar to each other, and the data within different categories are the least similar. It is expected that students who fall into different categories will be able to explain certain characteristics. Each student category number will be added to the student feature list as a new feature.

Second part: Prediction with LSTM neural network

After categorizing the students, it is time to create a NN on the data. In this step, training data is nourished to the LSTM neural network to use this data to produce a trained LSTM neural network. This LSTM neural network now can predict. Thus, in the test step, it will be capable to predict. This LSTM neural network can now be used as a classifier. By entering the training data into an LSTM neural network as well as the students 'academic performance, the NN is adjusted in such a way that it can now anticipate the students' performance based on the acquired characteristics. Because the LSTM neural network as a classifier allows training to be used for decision-making from now on.

3.3. Test Step

In the test step, the level of students' academic performance should be determined by entering the data. Therefore, first, the academic data of the incoming education student is assigned to one of the categories according to the centers of the categories obtained in the step of education. In fact, in this step, a group of students is selected who have the most similarities with the new student. It will lead to a more accurate judgment of new students. After this, the data of the new students are provided to the neural network LSTM. The neural network LSTM determines the academic performance of the new student based on the training given in the training step. This process leads to the incoming students being initially assigned to the category that most closely resembles that category. It makes decisions about incoming students more accurate. In the next step, the LSTM neural network, which is also trained in classification features, can comment on the new student's academic performance.

The training data: The training data is used for the training step and classification, but the test data is entered into the proposed flowchart test step section. According to the output of this step, the ability of the proposed method can be judged. Test data are entered into the test section one by one to be evaluated using the step test section approach and to determine their academic performance.

The academic data: The education data of the test students, which are selected from the main and random dataset, are hidden from the system before they enter the proposed system. After the system has made a decision about the performance of the students, this decision is compared with the value previously expressed and hidden in the dataset for this student to choose the quality of the proposed system.

3.4. Evaluate the proposed method

The current authors employed this method to test the efficiency of the suggested process using MATLAB programming language. In the results of this evaluation, the proposed method is called the Proposed Method, and the comparative method is called the Based Method. The experiments were conducted in an environment with the conditions detailed in Table 1.

Table 1. Test system specifications.

Item	Specifications
Processor	Intel Core 2 Due CPU T6570 @ 2.10GHz 2.10 GHz
Main memory	4.0 GB
Hard disk	320 GB
Operating system	Microsoft Windows 7 Ultimate
Programming language	Matlab R2019b

To evaluate categorical methods, four criteria are often used: True Positive, False Positive, True Negative, and False Negative, which are defined in the following of each of these criteria. According to these four parameters, the following four evaluation criteria are introduced:

$$\text{Precision} = TP/(TP+FP) \quad (1)$$

$$\text{Recall} = TP/(TP+FN) \quad (2)$$

$$\text{Accuracy} = TP+TN/(TP+TN+FP+FN) \quad (3)$$

$$F1=2\ \text{Precision}.\text{Recall}/(\text{Precision}+\text{Recall}) \quad (4)$$

4. THE PROPOSED METHOD

The dataset employed in the assessment [20] is a collection of educational data from the Kalboard 360 learning management system. Kalboard 360 has many components. learning management system developed to promote learning through the service of advanced technologies. Such a system allows simultaneous Users to get access to instructional resources through an Internet connection from any device. The dataset has 16 characteristics and 480 student files. These characteristics may be divided into three groups:

(1) Demographic characteristics such as gender and nationality.

(2) Characteristics of academic records such as educational stages and grade level.

(3) Behavioral characteristics like raising hands in class, opening resources, responding to parental surveys, and school satisfaction.

There are 175 women and 305 males in the sample. Numerous nationalities are represented among these students, including 179 from Kuwait, 172 from Jordan, 28 from Palestine, 22 from Iraq, 17 from Lebanon, 12 from Tunisia, 11 from Saudi Arabia, nine from Egypt, seven from Syria, six from the United States, Iran, and Libya, four from Morocco, and one from Venezuela. The dataset gathered over two semesters: In the first semester, 245 student files gathered, and in the second semester, 235 student files. Students separated into two groups depending on the number of absence days because the dataset includes the school attendance characteristic. 191 students with more than seven days off and 289 students with less than seven days off. This dataset includes a new set of features. It is a feature of parental delivery in the educational process. The feature of parental participation has the following two characteristics: examining parental response and school satisfaction. 270 parents responded to the survey, and 210 did not, 292 parents were satisfied with the school, and 188 did not. Eventually, students organized into three classes with marks L (grades 0 to 69), M (grades 70 to 89), and H (grades 90 to 100).

4.1. Evaluation Results

The evaluation results are presented in the form of two different scenarios. In the first scenario, the dataset is segmented, and instead the use of the entire dataset is only partially utilized in experiments.

In the second scenario, the size of the test and training dataset has altered. In the observing, we will examine these scenarios and the results received from their implementation in the form of evaluation parameters.

The First Scenario: As mentioned earlier, in the first scenario, experiments are conducted on the size of the dataset used. In this scenario, only 40% of the total dataset is randomly selected and used in experiments. In the next step, the size has reached 50%, and this process has continued until the entire dataset evaluated. It should be noted that in this scenario, the size of the training and testing set was equal to 70% and 30%. Precision, recall, accuracy, and F1 criteria will be used as evaluation criteria. The results of these evaluations illustrated in Figures 2 to 5.

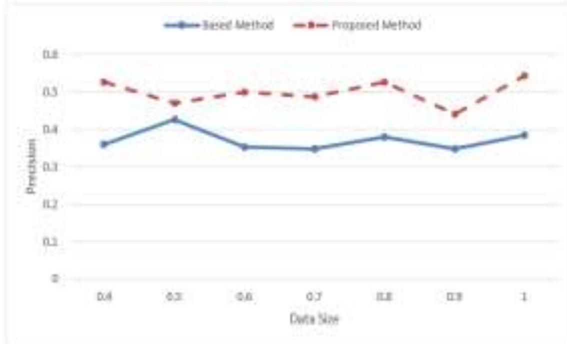

Fig. 2. Comparing Precision for different dataset sizes.

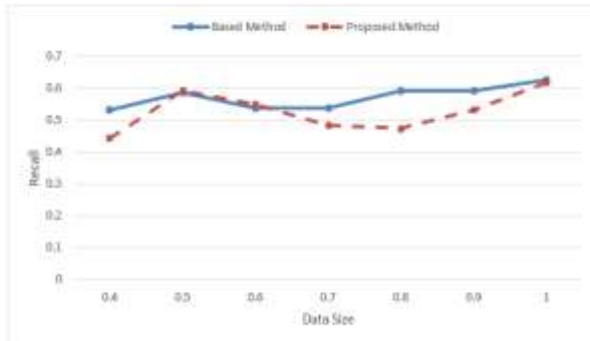

Fig. 3. Comparing Jump Recall for different dataset sizes.

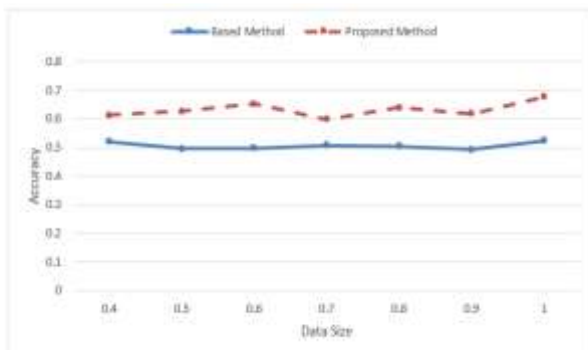

Fig. 4. Comparing Accuracy for different datasets.

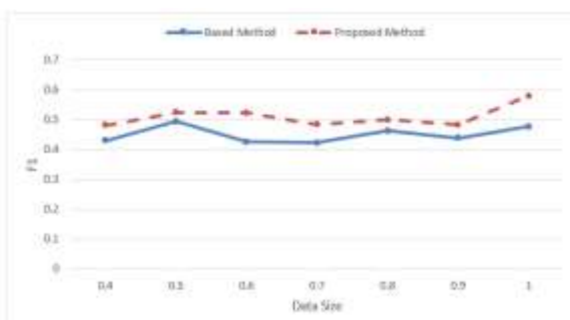

Fig. 5. Comparing F1 for different datasets.

As these results demonstrate, with the exception of recall, the proposed method in the precision, accuracy and F1 parameters has been able to perform better than the comparison method. According to the parameters used in the recall calculation, it can be seen that the proposed method only functioned worse in the FN parameter than the compared method. In fact, the proposed method has been proposed more than the method compared to the students in the same group, but in other parameters it has been able to show much better performance, and this can be seen in precision, accuracy and F1.

The Second Scenario: In this scenario, unlike the first scenario, the size of the dataset was constant and equivalent to the total dataset; but, it is the size of the test and training set has transformed. In fact, in this scenario, the size of the training set is initially considered to be 30% of the data, and of course, 70% is intended for testing. In the next case, the size of the training set is increased to 40% and the size of the test set is reduced to 60%. In the next step, the size of the training set is 50%, and this process continues until the size of the training set reaches 80%. The results of this scenario can be seen in Figures 6 to 9.

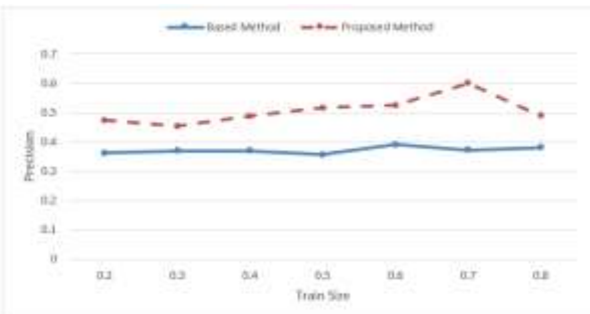

Fig. 6. Comparing Precision for different training set sizes.

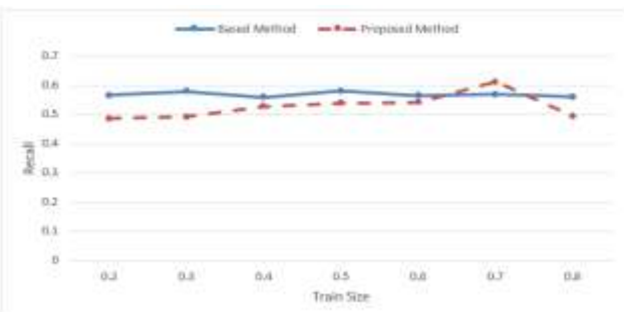

Fig. 7. Comparing Recall jumps for different training set sizes.

In the results of this scenario, as in the previous scenario, the proposed method, except for recall, in other cases has been able to provide better performance than the compared approach. As a result, it can be said that the proposed method only performed worse in the FN parameter than the compared method but, in other cases, it is the proposed method that delivers better results and leads to better precision, recall, and accuracy than the method compared.

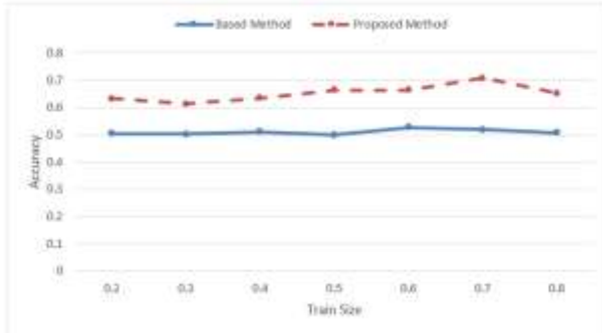

Fig. 8. Comparing Accuracy for different training set sizes.

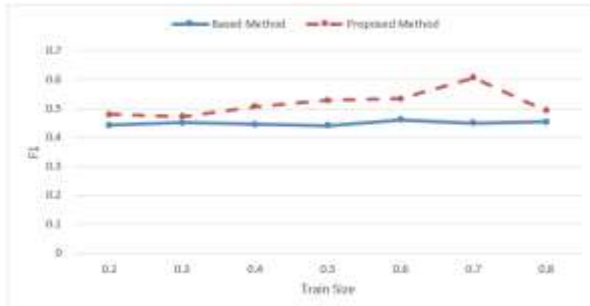

Fig. 9. Comparing F1 for different training set sizes.

5. CONCLUSION

Large amounts of data have been gathered as a result of recent developments in many different sectors, and this data is often kept in a variety of forms, including files, records, photographs, audio files, and videos. Although the data is employed in decision-making processes, the volume of data makes data administration and analysis difficult and complex. A suitable strategy for removing information from sizable repositories is needed in order to use data for better decision-making. Using data mining techniques, it is possible to extract useful knowledge from vast volumes of data. Data mining is a potent analytical technique that offers essential facts and information and can promote better decision-making. Nevertheless, because the educational data hierarchy differs from the classic data hierarchy, educational data mining algorithms are distinct from traditional data mining algorithms. Researchers have created novel methods for examining educational data in recent years, and this paper area has become autonomous.

In the current paper, to indicate student performance, we have operated an approach based on the classification algorithm and LSTM neural networks so that the process of student education can be well predicted.

MATLAB programming language has been used to estimate and compare the performance of the presented method. To compare, one of the latest algorithms was implemented along with the offered method. The results of the evaluations were performed in the form of two scenarios with different data sizes and different amounts of test and training data. For the evaluation, the dataset taken from an online educational system was used. The evaluation results, which are presented in the form of four well-known criteria of precision, recall, accuracy and F1, reveal that in the three evaluation criteria of precision, accuracy and F1, there is absolute superiority with the proposed method. But in the case of the recall criterion, the method compared has acquired better results. The use of classification systems such as support vector machines (SVMs) or NNs can equip acceptable results when the goal is to supply a predictive system. The current authors suggest to researchers interested in this field, to continue research in this field. In addition to using the mentioned categories, they can use classification methods along with these methods. Because, these algorithms can search for common features among users. These common features can have a positive effect on improving the quality of offers.

REFERENCES

[1] Baker, R.S., Corbett, A.T., Koedinger, K.R. “**Detecting Student Misuse of Intelligent Tutoring Systems**,” Proceedings of the 7th International Conference on Intelligent Tutoring Systems, pp.531-540, 2004.

[2] Romero, C., & Ventura, S.“**Data mining in Education**,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), pp.12–27, 2013.

[3] Araque, F., Roldan, C., & Salguero, A. “**Factors influencing university dropout rates**,” Journal of Computer & Education, 53, pp.563–574, 2009.

[4] Abu-Naser, S. S., Zaqout, I. S., Abu Ghosh, M., Atallah, R. R., & Alajrami, E. (2015). **Predicting student performance using artificial neural network: In the faculty of engineering and information technology**.

[5] Zacharis, Nick Z. "**Predicting student academic performance in blended learning using Artificial Neural Networks**." International Journal of Artificial Intelligence and Applications 7.5 (2016): 17-29.

[6] Sivagowry S, Dr.Durairaj M, “**PSO - An Intellectual Technique for Feature Reduction on Heart Malady Anticipation Data**”, International Journal of Advanced Research in computer science and software engineering, vol 4(9), pp 610-621, 2014.

[7] Zaidah ibrahim,” **predicting students’ academic performance: comparing artificial neural network, decision tree and linear regression**”, 21st annual sas malaysia forum, 5th september 2007.

[8] Mueen, A., Zafar, B., & Manzoor, U. (2016). **Modeling and predicting students' academic performance using data mining techniques.** International Journal of Modern Education and Computer Science, 8(11), 36.

[9] Hochreiter, Sepp, and Jürgen Schmidhuber. "**Long short-term memory**." Neural computation 9.8 (1997): 1735-1780.

[10] Zaidah ibrahim,” **predicting students’ academic performance: comparing artificial neural network, decision tree and linear regression**”, 21st annual sas malaysia forum, 5th september 2007.

[11] C. Romero, S. Ventura and E. García, “**Data mining in course management systems: Moodle case study and tutorial**”, Computers & Education, vol. 51, no. 1, (2008), pp. 368-384.

[12] P. M. Arsad, N. Buniyamin and J. L. A. Manan, “**A neural network students' performance prediction model (NNSPPM)**”, In Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on. IEEE, (2013), pp. 1-5.

[13] N. T. N. Hien and P. Haddawy, “**A decision support system for evaluating international student applications**”, In Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual.IEEE, (2007), pp. F2A-1.

[14] S. Abu Naser, “**Predicting Learners Performance Using Artificial Neural Networks in Linear Programming Intelligent Tutoring Systems**”, IJAIA, vol. 3, no. 2, (2012).

[15] G. Kanakana1, and A. Olanrewaju, “**Predicting student performance in Engineering Education using an artificial neural network at Tshwane university of technology**”, ISEM 2011 Proceedings, (2011) September 21-23, Stellenbosch, South Africa.

[16] E. Kyndt, M. Musso, E. Cascallar and F. Dochy, “**Predicting academic performance in higher education: Role of cognitive, learning and motivation**”, Earli Conference 2011, 14th edition, Exeter, UK, (2011).

[17] P. Mukta and A. Usha, “**A study of academic performance of business school graduates using neural network and statistical techniques**”, Expert Systems with Applications, Elsevier Ltd., vol. 36, no. 4, (2009).

[18] K. Stamos and V. Andreas, “**An Artificial Neural Network for Predicting Student Graduation Outcomes**”, Proceedings of the World Congress on Engineering and Computer Science (2008) “WCECS 2008”, San Francisco, USA.

[19] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). **Mining educational data to predict student’s academic performance using ensemble methods**. International Journal of Database Theory and Application, 9(8), 119-136.

[20] https://www.kaggle.com/aljarah/xAPI-Edu-Data.

[21] Sepp Hochreiter; Jürgen Schmidhuber (1997)."**Long short-term memory**".Neural Computation. 9 (8): 1735–1780.

[22] Li, Xiangang; Wu, Xihong (2014-10-15). "**Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition**". arXiv:1410.4281 [cs.CL].