Proceedings IFTE-2021, 0477-0489
doi:10.3897/ap.5.e0477

VII International Forum on Teacher Education

Research and forecasting of educational process parameters by using artificial intelligence

Anis F. Galimyanov* (a), Fail M. Gafarov (b), Liliya E. Khairullina (c), Alina I. Muzafarova(d)

(a), (b),(c),(d) Kazan Federal University, 420008, Kazan (Russia), 18 Kremlyovskaya street, anis_59@mail.ru

Abstract

In this paper we present the results of an interdisciplinary research based on the application of big data, date science, artificial intelligence and machine learning methods in educational analytics. Artificial intelligence techniques applied for the analysis of depersonalized data stored in the information and analytical system "E-education in the Republic of Tatarstan" from 2015 to 2020. BigData technologies were used in this work to perform high-performance computing related to initial preprocessing of raw data in computation cluster. By using the methods of artificial intelligence, we modelled one of the most important stages in the formation of the educational trajectories of schoolchildren, associated with the fact that after the 9th grade, schoolchildren either continue their studies in high school (grades 10-11), or move to the professional educational organizations. As the input data for neural network training, we used a vector containing the average marks for all quarters of pupils, obtained by using high-performance Dask-based cluster data processing system from initial raw data. We concluded that multi-layer neural network with two hidden layers was able to predict the pupil's pass to 10th grade, and achieved the best performance with classification accuracy exceeding 70%. Also, the performance of trained neural network had been analyzed by visualization of Receiver Operator Characteristic (ROC)-curve and by calculation of recall, precision, specificity and area covered by the ROC-curve (AUX) parameters.

Keywords: artificial intelligence, machine learning, neural networks, Big data, educational data mining, ROC-curve.

* Corresponding author. E-mail: anis_59@mail.ru

Introduction

In recent years, there has been an increase in the proportion of graduates of the 9th grade who leave school and continue further training in the system of secondary vocational education. So, if in 2017 proportion of 9th grade graduates who moved to the 10th grade in the Republic of Tatarstan was 47.7% [Lomteva & Bedareva, 2019], then in 2020 54% of ninth graders planned to continue their studies in secondary vocational educational institutions. The increase in the quantity of applicants to professional educational organizations after the 9th grade is a trend of recent years, not only in Republic of Tatarstan, but also in the whole of the Russian Federation. At the same time, the student's departure from school after the 9th grade does not always mean that he does not keep up with his studies. Often, going to college is a way to build a less stressful educational trajectory - a path to higher education, bypassing the USE, because college graduates can enter some universities based on the results of internal exams that universities and institutes conduct for them. However, maintaining a strong student body and their academic performance are important issues for schools and teachers. Predicting the continuation of a student's studies after the 9th grade would form an individual educational trajectory of the student, help teachers identify students who have the potential to successfully study in grades 10 and 11 and pass the Unified State Exam. On the other hand, such a prediction could be useful for pupils (and their parents) to assess future academic success based on their learning habits, work, and grades. This could help determine in a timely manner whether they should leave school after the 9th grade.

Currently, methods based on artificial neural networks (ANNs) are intensively used to solve the forecasting problem in various fields (Remus & O’Connor M,2001; Shiratori at all 2020) because ANNs can serve as a powerful and complex modelling tools for modelling nonlinear functions that often describes the real-world systems (Lau at all, 2019). Therefore, ANNs trained on historical educational datasets can be very useful for modelling the educational trajectories of schoolchildren. In this work we studied the possibilities of using neural networks (multi-layer feed forward neural network) in predicting the transition of a pupil from 9th to 10th grade on the basis of pupil’s mean marks for different subject.

Purpose and objectives of the study

Digitalization of the school educational process in Russia, associated with the introduction of electronic journals and diaries, has made educational data more accessible for analysis. The service "Electronic education" in the regions of Russia is the primary aggregator of educational data, the first layer of analytics of digital educational traces of students. On the basis of this level, educational data analysts, as a rule, make organizational and pedagogical decisions.

To date, there is a lack of publications on the analysis and interpretation of large datasets obtained from such services, especially with the use of artificial intelligence methods. This publication fills this gap to some extent.

The purpose of this work is to build an ANN based model for predicting the continuation of a schoolchildren study at the school after the 9th grade in the 10th grade. The data obtained from the system "Electronic education in the Republic of Tatarstan" for students of grades 6-10 for 2015-2020 are used as input parameters for neural network training. Note that this database contains more than two billion information units, including information about the academic performance of more than a million students and the professional activities of more than 120,000 teachers. The system contains anonymous data on 121,902 teachers, information on 90,741,876 lessons conducted and 1,034,312,802 grades. The main goal of this work is a study of machine learning methods possibilities and advantages for prediction the educational trajectory of schoolchildren, on the basis on the data available about their previous academic performance.

Literature review

In the last decade, a special field of education - Educational Data Mining (EDM) - has been actively developing. EDM means the direction associated with the extraction of information and knowledge about the educational process from large data sets (Big Data), in order to identify patterns encountered in it or formulate pedagogical theories (Romero & Ventura, 2010). Learning Analytics (LA), which measures, collects, analyzes, and presents data about students and the educational environment in order to optimize them, also has similar goals to the EDM direction. The number of researchers systematically dealing with EDM and LA issues is growing, and new works in these areas are published annually [Cruz-Jesus et al., 2020; Abu-Naser et al., 2015; Francis & Sasidhar, 2019, Isljamovic & Suknovic, 2014]. Among domestic researchers in this field, we can distinguish the works of (Kotova, 2019; Fiofanova, 2019; Belonozhko et al., 2017; Veryaev, 2016). It should be noted that most works in the field of EDM/LA are devoted to predicting student performance, identifying anomalous / extreme values in the education system, predicting exam success, etc. The following tools are used to achieve these goals: decision tree, artificial neural networks (ANN), k-nearest neighbor method, naive Bayesian classifiers, support vector machine, cluster algorithms (Isljamovic, 2014).

In EDM one of the most popular methods, are the methods, based on machine learning and artificial neural networks (Gafarov at all, 2020; Rastrollo-Guerrero at all, 2020).

A hybrid model that serves as the core design for a university admission recommender system, based on neural networks and decision tree classifier is resented in (Fong S., & Biuk-Aghai R. P., 2009). In this work the authors tested the system performance on the live data from sources of Macau secondary school students, and concluded that the system can be used to prediction suitable universities that match the students' profiles. In study by (Zabriskie at all, 2019) different machine learning methods (random forest and logistic regression models) were used to build early warning models of student success in introductory calculus-based mechanics and electricity and magnetism courses at a large eastern land-grant university serving approximately 30 000 students.

In another work (Mengash, 2020), authors have demonstrated, by using artificial neural networks, that early university performance of applicants can be predicted before admission, based on certain pre-admission criteria. The results, obtained by computer simulations showed that Scholastic Achievement Admission Test score is the pre-admission criterion that most accurately predicts future student performance. Authors also showed that ANN-based system has an accuracy rate above 79%, making it superior to other classification techniques (Support Vector Machines Decision Trees, and Naive Bayes). In the paper (Zhu at all, 2020) a model of university teachers' performance evaluation based on machine learning methods had been proposed. Authors proved through experiments the feasibility and practicability of the scientific research performance evaluation model based on artificial intelligence methods.

Application of machine learning methods for predicting students' performance is one of the most important topics for learning contexts such as schools and universities (Gafarov at all, 2020). In the review paper (Rastrollo-Guerrero at all, 2020), the authors provide a detailed overview of machine learning methods in EDM, based on the analysis of almost 70 scientific papers about different modern techniques. A neural network approach to classify student graduation status based upon selected academic, demographic, and other indicators presented in (Lesinski, at all 2016). Researchers developed a multi-layer feedforward network, and trained it on the data, obtained from institutional research databases. Several neural network architectures are compared by run time and different performance characteristics. Author concluded that a multi-layer neural network with 50 hidden neurons, hyperbolic tangent activation functions was able predict graduation success and achieved the best performance with classification accuracy exceeding 95% (Lesinski, at all 2016).

In conclusion of the analysis of the available literature, it should be noted that at present time, there are no publications regarding the application of machine learning methods and neural networks for the analysis of educational data (academic performance) of school students.

Most of the investigations are based on the analysis and prediction of higher educational institutions student's performance, which is due to the fact that the large and systematized educational datasets in general are available only in large universities.

Methodology

For the primary preprocessing of the initial raw data, which consisted in grouping data and calculating the average study quarter grades of pupils we developed a program module in Python programming language by using the Dask distributed computing framework (Rocklin, 2015). Dask is a flexible parallel big data processing system, designed to provide scalability and to extend the capabilities of existing Python packages and libraries. The main advantage of using the Dask system to solve our tasks is that this system allows parallel computations on data volumes that are larger than the available memory of single computer. For execution a high-performance calculation computing we deployed a cluster of 4 virtual machines (each VM 1TB HDD, 32 GB RAM, 16 CPU cores).

For predicting the transition of a pupil from 9th to 10th grade tasks we used multilayer feed-forward neural networks architecture (Ganesh & Abdesselam, 2003) by using Keras framework. We have reduced the task of predicting the transition of a pupil from 9th to 10th grade to a binary classification problem, and used the neural network with two hidden layers containing 50, 10 hidden neurons, respectively. To avoid overfitting, we regularize the network using dropout layers after each hidden layer, with dropout rate set to 20%. Dropout layer randomly prevents a proportion of neurons in hidden layer in each training loop from propagating their output into the next layer. The activation function for each hidden neuron is a rectified linear unit (ReLU) with dropout regularization of 20%. Since we are trying to classify a binary variable the output layer contains one neuron, corresponding to the probability of “passed to grade 10” and uses sigmiod activation. During the neural network training process weights are modified by minimizing the binary crossentropy loss function using the backpropagation algorithm, with adam optimizer.

In binary classification machine learning tasks, the output is either zero or one. For supervised learning task a confusion matrix (Susmaga, 2004)– a specific table layout for visualization of trained neural networks performance is used. It shows how well the model is performing on each class.

Table 1. Confusion matrix

Actual	**Predicted**	
	Positive	**Negative**
Positive	TP	FN

Negative	FP	TN

We evaluate the performance of trained neural network for binary classification by using four outcomes described in confusion matrix: true positive rate (TP), false positive rate (FP), true negative rate (TN), and false negative rate (FN). The four different classifications are:

- TP - the number of instances that are predicted positive and actually are positive. The higher value the better.

- FP - the number of instances that are predicted positive and actually are negative. The low the value the better.

- TN - the number of instances that are predicted negative and actually are negative. The higher the value the better

- FN - the number of instances that are predicted negative and actually are positive. The low the value the better.

From the confusion matrix many different statistics characterizing neural network's prediction accuracy can be computed (Goutte & Gaussier, 2005).

Accuracy is the fraction of correctly classified result, which is a very commonly can be calculated by the formula:

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Sensitivity (recall) is the is the fraction of positives that are correctly predicted as positive and is measured with:

Sensitivity =TP/(TP+FN)

For example, in medicine, the numerical values of sensitivity represent the probability of a diagnostic test identifies patients who do in fact have the disease. The higher the value the better.

Specificity is the proportion of the true negatives correctly identified by a neural network. This parameter suggests how good the neural network is at identifying negatives.

Specificity =TN/(TN+FP)

Precision is the ratio of the correctly positive classified by trained neural network to all positive labeled.

Precision=TP/(TP+FP)

F1-score conveys the balance between the recall and the precision, and it emphasizes the performance of a classifier on common and rare categories, respectively. It is measured with:

F1 score =2*(Sensitivity *Precision)/(Sensitivity *Precision)

Using F1-Measure, we can observe the effect of different kinds of data on a classification system (Goutte & Gaussier, 2005).

To present results for binary classification problems in machine learning Receiver Operator Characteristic (ROC) curves are commonly used (Saito & Rehmsmeier, 2015). This technique was originally developed to determine if a radar receiver were accurately detecting aircraft. The ROC curve is visualized by plotting the true positive rate against the false positive rate for all values of the decision threshold from 0 to 1. And therefore, ROC curves show how the number of correctly classified positive examples varies with the number of incorrectly classified negative examples. By using ROC curve, we can compare different models directly in general or for different thresholds, and the shape of the curve contains a lot of information false positive and false negative rate balance for different thresholds. By using ROC curve, we can to estimate a measure of the model's classification performance as measured by the area under the curve. The area covered by the curve (AUX) is the area between the ROC curve and the x-axis. In a model that is no better than guessing, the AUC will be 0.5 and the ROC curve will be a straight line. The bigger AUC values mean the better machine learning models is at different classes classification task, and maximum value for AUC is 1.

Results

By using high-performance Dask-based cluster data processing system we calculated average study quarter marks of 6,7,8,9 grade pupils for time interval from 2015 to 2020 years. As the input data of the neural network, we used a vector containing the average marks for all quarters of students in the following subjects: the Russian language, mathematics, literature, history, the English language, physical culture. We build a neural networks model for prediction school pupils who have passed from grade 9 to grade 10. This problem is considered as a machine learning binary classification problem.

The dataset is divided into two set of categories: “passed to grade 10” (positive) and “not passed to grade 10” (negative). Therefore, the target values for neural network training process are values from the set (0,1), 0- means that the pupil will non continue his teaching in the school at 10 grade, 1- means the opposite. The training data set was spitted into 70% training, 30% testing to facilitate model development, experimentation, and performance assessment. The network is trained for 100 epochs.

The confusion matrix for the classification results on train dataset is presented in Table 1. We can see that neural network better classifies actual negative data then the actual positive data, i.e., the neural network better predicts the pupils who does not continued his education in 10-th grade.

Table 2. Confusion matrix for trained neural network on test dataset.

Actual	**Predicted**	
	Positive	**Negative**
Positive	1433	1004
Negative	432	2038

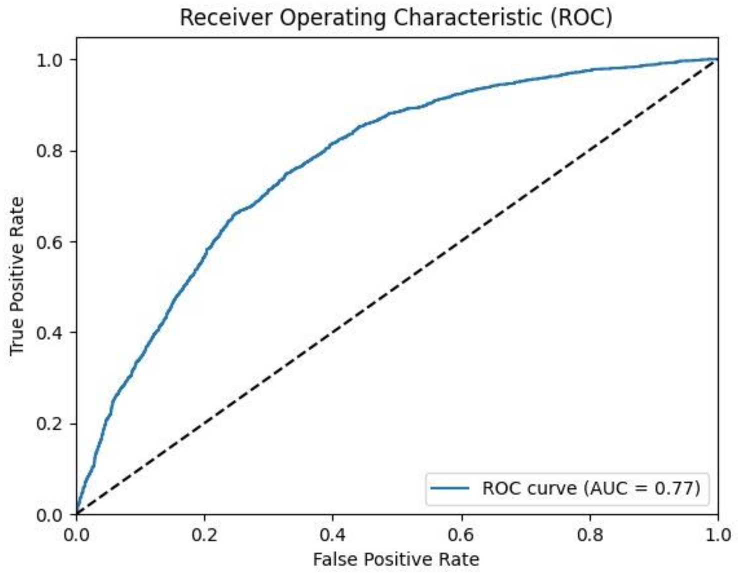
The performance metrics of trained neural network are presented in Table 3 shows that the trained neural network is able to classify the data with fairly good accuracy 71%.

Table 3. The characterizing of trained neural network’s prediction performance (Accuracy, Sensitivity, Specificity, Precision, F1-score) calculated from confusion matrix.

Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
0,71	0,83	0,59	0,59	0,69

The ROC curve for trained neural network is demonstrated in Figure 1. We measured the trained model performance also by using area under the ROC curve (AUC) evaluated on a test set. The value AUC=0.77 means that the neural network model has the potential for forecasting the pupil’s educational trajectory.

Figure 1. ROC curve for trained neural network.



Discussion

The 9th grade boundary has a significant impact, on the one hand, on the motivation of schoolchildren, and on the other hand, it seriously affects teacher assessment practices. On the whole, this makes a strict picture of the educational trajectories formed in the Russian school. The decision to continue education in the 10th grade is motivated not only by the academic performance of the students, but also by the parents and the social environment. Therefore, for a full and multifaceted study and modeling of these processes, complex models based on the methods of artificial intelligence and machine learning are needed.

By now, there are many different national databases containing detailed information on the educational process. The use of artificial intelligence methods and neural networks currently in educational analytics opens up new possibilities in predicting the educational trajectories of high school students.

In the future, it is planned to continue the study of this issue in the context of the municipal districts of the Republic of Tatarstan.

Conclusion

In this work, we have demonstrated the potential of using neural networks in predicting the transition of a pupil from 9^{th} to 10th grade.

The prediction is based on the using of the multi-layer feed forward neural network on pupil’s mean marks for different subject. The values of the trained neural network performance parameters (Accuracy, Sensitivity, Specificity, Precision, F1-score) suggests that the neural network was able to identify some patterns in educational process. We are confident that the better results can be obtained if more parameters characterizing the educational environment and the educational process will be used for training neural networks. It should be noted, that the approach developed by us and the system are easily expandable, if a new data on the educational process will appear, we can easily add these new variables to our model, what can significantly increase its predictive accuracy. The results obtained in this work can be useful for employees of the education system, heads of educational organizations, and researchers.

Acknowledgements

The reported study was funded by RFBR, project number 19-29-14082

References

1. Abu-Naser, S., Zaqout, I., & Abu Ghosh, M., & Atallah, R., & Alajrami, E. (2015). Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology. *International Journal of Hybrid Information Technology*. *8*. 221-228.

2. Belonozhko, P.P., Karpenko, A.P., & Khramov, D.A. (2017). Analysis of educational data: directions and prospects of application. *Internet magazine "Science"*. *4*(9). 1-21. Available at: http://naukovedenie.ru/PDF/15TVN417.pdf

3. Cruz-Jesus, F., Castelli, M., & Oliveira, T. & Mendes, R. & Nunes, C. & Sa-Velho, M. & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country, *Heliyon*, *6*, 1-11.

4. Fiofanova, O. A. (2020). Big data analysis in education: methodology and technology. Moscow: Delo.

5. Fong S., & Biuk-Aghai R. P (2009). An Automated University Admission Recommender System for Secondary School Students, *6th International Conference on Information Technology and Applications, ICITA 2009*, 37-42

6. Francis, B., & Sasidhar Babu, D. (2019). Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems. 43*. 1-15.

7. Ganesh, A. & Abdesselam, B. (2003). A generalized feedforward neural network architecture for classification and regression. *Neural Networks*. 16, 5–6, 561–568. Available at: https://doi.org/10.1016/S0893-6080(03)00116-3

8. Gafarov F.M., Rudneva Ya.B., Sharifov U.Yu., Trofimova A.V., Bormotov P.M. (2020). Analysis of Students' Academic Performance by Using Machine Learning Tools. *Proceedings of the International Scientific Conference "Digitalization of Education: History, Trends and Prospects" (DETP 2020),* 570-575. Available at: https://www.atlantis-press.com/proceedings/detp-20/125940239

9. Goutte, C. & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and *F*-Score, with Implication for Evaluation. In: Losada D.E., Fernández-Luna J.M. (eds) Advances in Information Retrieval. ECIR 2005. *Lecture Notes in Computer Science*, 3408, 345-359. Available at: https://doi.org/10.1007/978-3-540-31865-1_25

10. Isljamovic, S., & Suknovic, M. (2014). Predicting students' academic performance using artificial neural network : a case study from faculty of organizational sciences. *ICEMST 2014: International Conference on Education in Mathematics, Science & Technology / The Eurasia Proceedings of Educational & Social Sciences (EPESS*), *1*, 68-72. Available at: https://dergipark.org.tr/en/download/article-file/332072

11. Kotova, E. E. Forecasting the success of learning in an integrated educational environment using online analytics tools. *Computer tools in education. 4*. 55-80.

12. Lau, E.T., Sun, L. & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Appl. Sci.* 1, 982. Available at: https://doi.org/10.1007/s42452-019-0884-7

13. Lesinski G., Corns S., & Dagli C. (2016) Application of an Artificial Neural Network to Predict Graduation Success at the United States Military Academy, *Procedia Computer Science*, 95, 375-382.

14. Lomteva, E.E, & Bedareva, L.Yu. (2019). Transition of 9th grade school graduates to the vocational education system: problems and prospects. *Domestic and foreign pedagogy*, *3* (60), 56-71. Available at: https://cyberleninka.ru/article/n/perehod-vypusknikov-9-h-klassov-shkol-v-sistemu-professionalnogo-obrazovaniya-problemy-i-perspektivy1.

15. Mengash H. A. (2020). Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. *IEEE Access,* 8, 55462-55470. Available at: doi: 10.1109/ACCESS.2020.2981905.

16. Rastrollo-Guerrero JL, Gómez-Pulido JA, & Durán-Domínguez A. (2020) Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Applied Sciences*. 10(3):1042. Available at: https://doi.org/10.3390/app10031042

17. Remus W., & O'Connor M. (2001) Neural Networks for Time-Series Forecasting. In: Armstrong J.S. (eds) Principles of Forecasting. *International Series in Operations Research & Management Science,* 30. Available at: https://doi.org/10.1007/978-0-306-47630-3_12

18. Rocklin, M. (2015) Dask: Parallel Computation with Blocked algorithms and Task Scheduling, *Proceedings of the 14th Python in Science Conference*, 130 - 136. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.825.5314&rep=rep1&type=pdf

19. Romero C., & Ventura S. (2010) Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews)*. 2010, *6*(40), 601-618.

20. Saito, T & Rehmsmeier, M (2015) The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE* 10(3): e0118432. Available at: https://doi.org/10.1371/journal.pone.0118432

21. Shiratori T, Kobayashi K, & Takano Y (2020) Prediction of hierarchical time series using structured regularization and its application to artificial neural networks. *PLoS ONE* 15(11): e0242099. Available at: https://doi.org/10.1371/journal.pone.0242099

22. Susmaga, R. (2004) Confusion Matrix Visualization. In: Kłopotek M.A., Wierzchoń S.T., Trojanowski K. (eds) *Intelligent Information Processing and Web Mining. Advances in Soft Computing*, vol 25. Springer, Berlin, Heidelberg. Available at: https://doi.org/10.1007/978-3-540-39985-8_12

23. Veryaev A.A., & Tatarnikova G.V. (2016). Ducational data mining and learning analytics - directions of development of educational qualitology. *Teacher of the XXI century, 2*, 150-160. Available at: https://cyberleninka.ru/article/n/educational-data-mining-i-learning-analytics-napravleniya-razvitiya-obrazovatelnoy-kvalitologii.

24. Zabriskie C., Yang J., DeVore S., & Stewart J. (2019). Using machine learning to predict physics course outcomes. *Phys. Rev. Phys. Educ. Res*. 15, 020120.

25. Zhu Z., Dai W., Wang J., Li J., & Hu P. (2020) Application of Machine Learning in the Evaluation Model of Scientific Research Performance of Teachers. Big Data Analytics for Cyber-Physical System in Smart City. BDCPS 2019. *Advances in Intelligent Systems and Computing*, 1117, 260-270.