Modeling educational data in Afghanistan using machine learning methods

Azizi Bahram
Independent researcher
ITMO National Research University
Saint Petersburg, Russia
bahramuddinazizi@gmail.com
ORCID 0000-0000-0000-0000

Olga B. Egorova
Candidate of Philological Sciences
Associate Professor
ITMO National Research University
Saint Petersburg, Russia
oegorova@itmo.ru
ORCID 0000-0002-5967-5638

Received 01.12.2024
Accepted 24.01.2025
Published 28.02.2025

UDC 004.428.092.8:371.491.1(945.53)
DOI 10.25726/p0203-5421-9376-e
EDN BTOKYQ
VAK 5.8.1. General pedagogy, history of pedagogy and education (pedagogical sciences)
OECD 05.03.HA. EDUCATION & EDUCATIONAL RESEARCH

Abstract

This paper examines various factors affecting a student's academic performance at the stage of secondary education in Afghanistan. Of course, academic achievements play a crucial role in determining a student's future profession, as well as their goals in an academic or professional environment. Academic success is influenced not only by academic performance itself, but also by other factors, including individual differences. In this work, various machine learning algorithms were used to analyze the factors influencing a student's grades in the last grade of school, in particular, learning with a teacher (regression and classification) and without a teacher (clustering algorithms). The purpose of the study was to determine which factors significantly affect student performance for the entire sample under consideration, which factors are more individual in nature, as well as the degree of their influence. The results showed that various machine learning algorithms can be successfully used to determine the factors influencing students' academic success. Students' final grades are undoubtedly influenced by the results of their studies in previous years, as well as the time spent on learning. But it is very important to note that the family situation and the balance between study and leisure activities also affect the student's success. In particular, the level of education and the mother's job have a great impact on the student's academic success: if the student's mother is educated and, especially, if she works as a teacher, this significantly increases the student's chances of academic success.

Keywords

educational data modeling, academic success factors, machine learning methods, neural networks, secondary education data, educational data analytics, intellectual analysis of educational data.

Introduction

In every country and, in particularly, in Afghanistan, high school years are very important, as the grades of these years determine a student’s performance in school and their success in further university entrance examinations. Academic success is of great importance as it can lead to better job opportunities and higher salaries in future. In addition to school grades, other factors that can affect academic success include attendance, participation in class discussions, and assignment completion. By identifying these factors and eliminating negative influences, students can improve their academic performance and achieve their goals.

The education system in Afghanistan is very poor and classic, with almost no usage of technology. The analytical techniques that are being used for improving the education system are outdated. This research aims to address these issues by using artificial intelligence and machine learning algorithms to determine the factors that affect academic success and their degree of influence in the context of Afghanistan’s education system. The understanding of underlying reasons provides valuable insights into the dynamics of academic achievement and offer a foundation for future research and policy-making in the field of education in Afghanistan, enabling educators, policymakers, and stakeholders to make informed decisions and strategies to enhance academic achievement.

The core challenge of this study was to evaluate the impact of various factors on academic achievement and subsequently predict academic success using appropriate artificial intelligence algorithms. Moreover, the predictive models were built using suitable machine learning algorithms that can effectively handle the complexity and diversity of the influencing factors.

Materials and research methods

Let’s consider what factors are mentioned in the studies that can influence the students’ academic achievement.

One of the important factors is the students’ attitude toward the languages, both mother tongue and foreign languages (Kang, 2021; Smith, 2023). The more students master their own native language, the easier it will be to learn a foreign language (Sidole, 2019). The proficiency in the mother tongue and the passion towards studying a foreign language is often connected with general attitude to studying.

The inclusion of a student's demographic information in academic studies is crucial for defining and categorizing student profiles and academic success, as different studies show (Ghaleb,2021).

The impact of cultural differences on the academic achievements varies significantly across different cultures and countries (Gebauer, 2021; Wang, 2020). The quality and effectiveness of education are significantly influenced by a country's social and economic welfare.

Gender differences in academic achievement, particularly in mathematics, have been a subject of extensive research. Some studies suggest that male students tend to approach problem-solving tasks with greater speed and accuracy. However, this is not a universal finding and can vary based on numerous factors, including the specific educational context and cultural background (Perez, 2021).

Student’s involvement in social activities during their educational journey generally enhances their academic performance. A student who participates in these activities and effectively manages their time is likely to see an improvement in their academic success (Xerri, 2018). For instance, students engaged in sports often exhibit discipline, which they apply across various areas of their life.

Humans, being social creatures, have various emotional needs. When these needs are unmet, they may experience negative feelings such as anxiety and loneliness. Social support plays a crucial role in mitigating these feelings. In today's competitive environment, factors like the continuity of education, social relationships, economic conditions, and more necessitate social support for a healthy life. This support can come from family, friends, teachers, specialists, and others in our social network (Cherry, 2023).

Results and discussion

One of the factors influencing students' academic success is their motivation towards their lessons. The social support received from teachers, in particular, has a positive impact on students' academic performance (Lei, 2018).

Effective and planned studies lead to permanent learning that goes beyond rote memorization and reaches the level of cognition and above, resulting in academic achievements (How to study effectively: 12 secrets for success, 2017). The motivation to study a course is driven by the purpose of the course, the area of knowledge it covers, and how this knowledge will benefit the student.

A student's knowledge of efficient study techniques is another key factor influencing academic success. While each lesson has its own unique study techniques, students also have individual learning methods (How to study effectively: 12 secrets for success, 2017).

Time management is a crucial factor in studying efficiency. Factors such as unconscious internet use and unnecessary time spent commuting can prevent students from studying and lead to time losses. To mitigate these losses, students should plan their time and adhere to this plan (Wolters, 2021).

Emotional intelligence also plays a crucial role in academic success when managed correctly (Why you need emotional intelligence to succeed at school, 2020). Life satisfaction, another affective element, is intertwined with emotional intelligence, and these variables influence each other reciprocally (Kong, 2019). This interplay involves the student's overall perspective on life, their ability to define life positively, their zest for living, their level of life satisfaction, and the meaning they attribute to life. Consequently, life satisfaction, emotional intelligence, and academic achievement are interconnected factors that mutually influence each other.

For instance, a student experiencing low academic success may see a decrease in life satisfaction and suffer emotional distress (Kristensen, 2023). Conversely, during emotionally challenging times, a decrease in life satisfaction can lead to a drop in academic achievement. These variables dynamically and directly impact each other (Ge, 2021).

As technology advances, computer science is playing a pivotal role in understanding the elements that influence academic success. Machine learning algorithms are employed to model training data, executing classification, estimation, and categorization studies by integrating data mining and artificial intelligence. Educational data mining has become an effective tool for uncovering hidden relationships in educational data and predicting students' academic achievements. Let's consider what machine learning algorithms are used in educational data mining to find the factors impacting the educational effectiveness and to predict and increase the student effectiveness.

One study conducted in Malaysia compared the use of an artificial neural network (ANN) technique and a combination of decision tree classification and clustering to predict and classify students' academic performance. The data used in this study were taken from the National Defense University of Malaysia and included information on 85 students. This data contained demographic information such as gender and race, educational information such as computer skills and courses taken, and personality information such as interests and motivation (Wook, 2009).

Another research paper compared the accuracy of decision tree and Bayesian network algorithms in predicting the academic performance of undergraduate and postgraduate students at two very different academic institutes: Can Tho University (CTU) in Vietnam and the Asian Institute of Technology (AIT) in Thailand. Despite the differences in student populations at these two institutes, the data-mining tools achieved similar levels of accuracy in predicting student performance: 73/71% for (fail, fair, good, very good) and 94/93% for (fail, pass) at CTU/AIT respectively. These predictions were most useful for identifying failing students at CTU (64% accurate) and selecting very good students for scholarships at AIT (82% accurate). The decision tree was consistently 3-12% more accurate than the Bayesian network (Pojon, 2017).

Creation of student achievement prediction models to predict student performance in academic institutions is another area of research in data mining techniques. One study proposed a prediction system that used students' 10th and 12th grade marks and previous semester marks to predict their final grades. The study evaluated the use of binomial logistic regression, decision tree, entropy, and KNN classifiers. This system could help students identify their final grade and improve their academic behavior, in order to achieve higher scores (Altabrawee, 20190.

In the study of Turkish students, various machine learning algorithms were used to predict the final exam grades of undergraduate students using their midterm exam grades as source data. Such algorithms as random forests, nearest neighbor, support vector machines, logistic regression, Naïve Bayes, and k-nearest neighbor were built and compared. The results show that the proposed model achieved a classification accuracy of 70-75% using only three types of parameters: midterm exam grades, department data, and faculty data (Yağcı, 2022).

Another research work investigates the application of six machine learning algorithms to student performance prediction, using datasets made up of general students' information available at the Atilim University administrative systems only. The study also investigates whether the number of courses predicted together is directly or inversely proportional to the performance of the classifiers used. Naïve Bayes and K-nearest neighbors algorithms proved to be the best among the compared algorithms. The results also showed that as the number of courses being predicted together increases, the prediction performance decreases. Data preprocessing generally improves the performance of the machine learning algorithms (Vijayalakshmi, 2019).

One more work aimed to develop an academic prediction model for the student for the UCI (University of California Irvine). The internal marks were used for the prediction that are a combination of attendance marks, average marks from two examinations and marks of assignment. The teachers can therefore classify students and start predicting their performance at an early stage. Better results can be expected in the final examinations due to early predictions and solutions (Al-Juboori, 2019).

In this study the following features reflecting students' demographic structure, socioeconomic status, health/sports/activity characteristics, attitude, and academic success were analyzed: school, sex, age, address, family size, parental status, parental education, travel to school time, study time, failures, school support, family support, paid activities, nursery, higher education aspirations, internet access, romantic relationships, family relationships, free time, going out, weekday alcohol consumption, weekend alcohol consumption, health, absences, and grades in 10th, 11th, and 12th grades. Additionally, it considers the mother's and father's jobs, the guardian's identity, and the reason for choosing the school.

These factors were pinpointed through a survey given to Afghan students. This research was based on the responses of high school students from two schools in Afghanistan. Therefore, it would be worth to investigate the situation in other regions or educational contexts of Afghanistan to be able generalizable the findings. The dataset consists of 33 features with 16 numerical and 17 categorical features. More than 1000 Afghan students participated in the survey study.

In cases where supervised machine learning algorithms were used, the marks of 10^{th}, 11^{th}, 12^{th} grades, and combined averages were taken as target variables. In other cases the target variable was the 12^{th} grade and other features were used as predictors. There was no missing data in the dataset. The features in the dataset can be extracted, compressed, or categorized by applying Scale, Feature selection, LDA, PCA methods, provided that they change with the algorithms used.

Various regression models were used to predict the academic achievement including Linear Regression, Lasso, Ridge, K-Nearest Neighbors (KNN), Polynomial Regression, Decision Tree, Random Forest, and Support Vector Regression (SVR) both with and without hyperparameter tuning.

If we consider the models without hyperparameter tuning (i.e., grid search was not used), the Random Forest model achieved the highest R2 score of 0.86, indicating it was the most effective at predicting student performance. The Decision Tree and Support Vector Regression (SVR) models had lower R2 scores, suggesting they might not be the best fit for this particular dataset (Table 1).

With hyperparameter tuning subtle improvements were achieved in terms of evaluation scores. The best model is again Random Forest with about 0.87 R2 score. Furthermore, the other models are also having a decimal improvement.

Table 1. Regression models results

	Model	R2 Score	MAE	MSE	RMSE	Avg R2 Cross-Val
1	Linear Regression	0,83	5,33	68,93	8,30	0,81
2	Lasso	0,85	4,75	61,31	7,83	0,82
3	Ridge	0,83	5,28	67,98	8,24	0,81
4	K-Nearest Neighbors	0,83	5,07	67,60	8,22	0,80

5	Decision Tree	0,85	4,68	58,98	7,68	0,83
6	Random Forest	0,87	4,48	51,05	7,14	0,83
7	Support Vector Regression	0,78	5,36	87,28	9,34	0,74
8	Polynomial Regression with Linear Regression	0,83	5,33	68,93	8,30	0,81

Various classification models were evaluated including Logistic Regression, Decision Tree, Random Forest, SVM, KNN with and without hyperparameter tuning in two kinds of classification binary and multi-class classification. We transform our data to such a kind that it is possible to produce a binary classification. And converted the column 12th according the Table 2.

Table 2. Data conversion for binary classification

12^{th} grade	Class
>=45 (passed)	1
<45 (not passed)	2

Without hyperparameter tuning, the best results were obtained using the SVM model and the KNN model.

Table 3. Results of binary classification

	Model	Recall	Precision	F1	Accuracy	Roc Auc	Log Loss
1	Logistic Regression	0.93	0.95	0.94	0.90	0.95	0.25
2	Decision Tree	0.93	0.96	0.94	0.90	0.84	0.18
3	Random Forest	0.95	0.95	0.95	0.91	0.94	0.21
4	SVM	0.96	0.95	0.95	0.92	0.97	0.18
5	KNN	0.96	0.95	0.96	0.92	0.94	0.70

With hyperparameter tuning, we can observe, that SVM, Random Forest, and KNN are having higher performance. In summary, if we consider all metrics, the SVM model with tuned parameters appears to provide the best overall performance with F1 score equal to 0.96 (Table 3). The data was preprocessed to such a kind that a multi-class classification was applied. The 12-th grade was converted as shown in the Table 4.

Table 4. Data conversion for multiclass classification

12^{th} grade	Class
>=85	A
>=70 & <85	B
>= 55 & <70	C
<55	D

It seems the best models for multi-class classification both with or without the hyperparameter tuning were Logistic Regression and Random Forest with about 0.75 evaluation scores (Table 5).

Table 5. Multicalss classification results

#	Model	Recall	Precision	F1	Accuracy
1	Logistic Regression	0,77	0,76	0,77	0,77
2	Decision Tree	0,75	0,75	0,75	0,75
3	Random Forest	0,76	0,76	0,75	0,76
4	SVM	0,68	0,68	0,68	0,68
5	KNN	0,52	0,55	0,52	0,52

In this section, various ensemble models were evaluated including Ada Boost, Xgboost and Bagging both with and without hyperparameter tuning to optimize their performance.

First, let's discuss ensemble regression models. The ensemble models without hyperparameter tuning provided decent results, but there was room for improvement. The models managed to capture the general trend in the data but struggled with some of the more complex patterns. Overall, Gradient Boosting appears most effective, with AdaBoost and Bagging following respectively. High MSE and RMSE values for all models suggest sensitivity to outliers.

After tuning the hyperparameters, the performance of the ensemble models improved significantly. They were able to better capture the complexity in the data, leading to more accurate predictions. The best results shows Ada Boost model (Table 6).

Table 6. Ensemble regression models results

	Model	R2 Score	MAE	MSE	RMSE	Avg R2 Cross-Val
1	Bagging	0,8617	4,776	55,3118	7,4372	0,8193
2	AdaBoost	0,862	4,7181	55,1869	7,4288	0,8194
3	Gradient Boosting	0,8612	4,7349	55,4994	7,4498	0,8208

For binary classification tasks, we used ensemble models to categorize data into two classes as we did in classification models. The ensemble models without Hyperparameter Tuning performed reasonably well, correctly classifying a large proportion of the instances. However, they had some difficulty with instances near the decision boundary. Overall, Bagging appears to perform best in terms of recall and precision, with XGBoost and AdaBoost following respectively.

Tuning the hyperparameters improved the models' ability to correctly classify instances near the decision boundary. This led to an increase in the overall classification accuracy. Overall, AdaBoost appears to perform best in terms of recall and precision, with XGBoost and Bagging following respectively (Table 7).

Table 7. Ensemble models and binary classification results

#	Model	Recall	Precision	F1	Accuracy	Roc Auc	Log Loss
1	Bagging	0,9296	0,9691	0,949	0,914	0,9645	0,3943
2	AdaBoost	0,9333	0,973	0,9527	0,9204	0,9519	0,198
3	Gradient Boosting	0,937	0,9656	0,9511	0,9172	0,9626	0,2238

For multi-class classification tasks, we used ensemble models to categorize data into more than two classes as we did in the Classification Models. The ensemble models without Hyperparameter Tuning showed promising results, correctly classifying many instances. However, they struggled with some of the classes that were less represented in the data. Among the three classifiers (Bagging, AdaBoost, and XGBoost), the XGBoost model is performing particularly well. It achieves an accuracy close to 81.2%, which is the highest among the three. Additionally, its precision, recall, and F1-score are also impressive. Overall, XGBoost appears to be the most effective model for this multi-classification task.

After tuning the hyperparameters, the ensemble models were able to better handle the classes that were less represented in the data, leading to a more balanced classification performance across all classes. Overall, XGBoost appears to perform best in terms of recall and precision, with Bagging and AdaBoost following respectively (Table 8).

Table 8. Ensemble models and multiclass classification results

#	Model	Recall	Precision	F1	Accuracy
1	Bagging	0,8	0,8	0,8	0,8
2	AdaBoost	0,68	0,68	0,65	0,68
3	XGBoost	0,81	0,81	0,81	0,81

Clustering models like K-Means and Hierarchical Clustering were used to group students based on their features. The clusters helped us understand the common characteristics of students who have similar academic performance. The results of both models were similar.

Balanced students:

– These students have a balanced profile with moderate values across most features;
– They have a higher level of maternal education (Medu) and a higher study time;
– They also have a higher proportion choosing a school for its reputation4
– Interpretation: Cluster 0 represents students who maintain a balanced approach to academics. They are neither high-achievers nor struggling academically.

Struggling students:

– These students seem to be struggling, with lower values for school support, family support, and 12th-grade performance.
– They have a higher travel time, indicating they might live further from school.
– They also have a higher proportion choosing a school for its courses.

So, cluster 1 likely includes students who face challenges in their academic journey. They may need additional support and resources.

Average performers:

– These students have average values across most features;
– They have a higher proportion of fathers with jobs in services;
– They also have a higher proportion choosing a school for its courses.

That is, cluster 2 represents students who consistently perform at an average level academically.

High support students:

– These students have high family support and paid extra classes;
– They have a higher proportion of mothers with jobs at home and fathers with other jobs;
– They also have a higher proportion choosing a school for its content.

That means that cluster 3 includes students who strike a balance between academics and other aspects of life. They achieve reasonably well without extreme pressure.

High achievers:

– These students have high values for maternal education, study time, and 12th grade performance;
– They have a higher proportion of mothers with jobs as teachers and fathers with jobs in services.
– They also have a higher proportion choosing a school for its reputation;

So, cluster 4 represents high-achieving students who excel academically, likely due to strong parental support and personal motivation.

Keras Regressor and Keras Classifier were used to capture the non-linear relationships between features. With a suitable number of hidden layers and neurons, ANN models achieved good performance in predicting student academic performance: R2 score is approximately 0.72, which indicates that about 72% of the variance in the target variable can be explained by the ANN model (Table 9).

Table 9. ANN results

ANN	
Mean Squared Error (MSE)	99,43
Root Mean Squared Error (RMSE)	9,97
Mean Absolute Error (MAE)	7,24

R-squared (R2)	0,72

Overall, the ANN model shows a reasonable performance in the regression task, with a relatively high R2 score. However, the errors (MSE, RMSE, and MAE) suggest there may be room for improvement. We can observe that the model is performing really well on Binary Classification with reasonably good values of matrix (Table 10).

Table 10. Binary Classification with ANN results

Binary Classification with ANN	
Accuracy	0.917
Precision	0.951
Recall	0.951
F1 Score	0.951

Multi Class Classification results are not as good as the one we had for Binary Classification (Table 11).

Table 11. Performance Metrics of the Multi-Class Classification Model with ANN

Multi-Class Classification with ANN	
Accuracy	0.828
Precision	0.607
Recall	0.624
F1 Score	0.615

We could add more layers to the network, but it could make the model more prone to overfitting. Besides, all models provided valuable insights into the factors affecting student academic performance. However, the choice of model depends on the specific requirements of the task, such as the interpretability of the model, the trade-off between bias and variance, and the computational efficiency.

Firstly, we analyzed the most important factors why students select a particular school for their studies are the closeness to the home, the course importance for the future and the school reputation. The following figure illustrates the correlation between the target variable and the other features.

The most important features are according the correlation of the features. According to this chart, you can see that the final score is not strongly correlated with any other feature except the 11th grade (0.92). But other features like family attitude towards studying, the family background (parental education) and jobs (educational), the possibility to find time to study are also important for the educational results.

In this section, we will showcase the results of our models, focusing on the importance of features that were crucial for predicting the target variable during the training process. AdaBoosting and Random Forest: The 11th grade is the most important feature, standing out significantly. This is logical as a student's performance in the 11th grade is likely to influence their final year score. However, it's important to note that this doesn't diminish the importance of earlier grades. The foundation built in earlier years can significantly impact a student's ability to perform well in the 11th grade. In addition, other important features include student's absences, the study time, and 10th grade. These factors directly relate to a student's academic engagement. Regular attendance, dedicated study time, and performance in the 10th grade can provide a strong foundation for success in the 11th grade. Furthermore, for Ada Boosting the following factors are important: the student's age, the travel time from home, relations within the family, and even the going out time are also showing subtle importance.

For random forest 11th grade feature also dominates, but other features such as student's age, the amount of free time, the health condition, even the going out time also play crucial role. These features represent aspects of a student's personal and social life, which can indirectly affect academic performance. For instance, maintaining good health and balancing free time activities with study time are essential for overall well-being and academic success. The Artificial Neural Network (ANN) model has identified the 11th grade, parents' job, and reasons for choosing a school the most important features for prediction. The 11th grade significantly outweighs the others in terms of importance, as it is expected and we have already discussed this in the previous models. The Random Forest and Gradient Boosting models perform similarly, with an R2 score of 0.87, indicating that they explain 87% of the variability in the target variable.

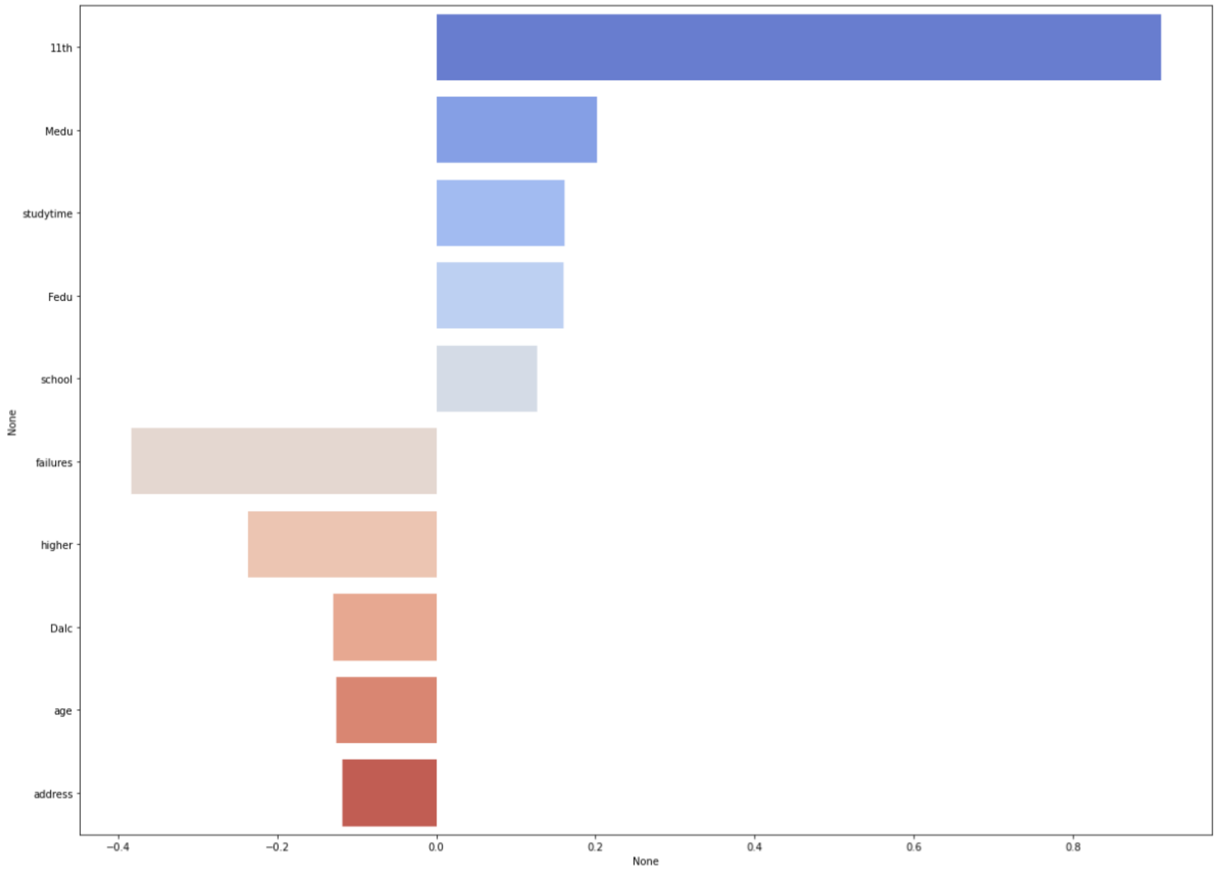

Figure 1. Correlation of the features and the 12th grade. Designations: 11th – marks in the 11th grade, Medu – mother's education, studytime – time spent on studies, Fedu – father's education, school – school name where does a student study, failures – the number of classes in which students had failures in the past, higher – means the plans of a student to have further higher education, Dalc – means drugs consumption during the weekdays, age – is the student's age, address – means urban or rural living.

However, the Random Forest model has a slightly lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), suggesting it might be a better choice for this dataset. The Artificial Neural Network (ANN) model has a significantly lower R2 score of 0.72, indicating it's less effective at predicting the target variable in this case. All three models – SVM, AdaBoost, and ANN – show high performance, with F1 scores and accuracy around 0.95 and 0.92 respectively. This suggests that they are all effective at classifying the data into two classes. The SVM model has the highest Roc Auc score of 0.97, indicating it has the best trade-off between sensitivity and specificity among the three models.

The XGBoost model performs the best among the three models, with a recall, precision, F1 score, and accuracy all at 0.81. This indicates that it's effective at classifying the data into multiple classes. The ANN model has the highest accuracy of 0.83, but its recall, precision, and F1 score are significantly lower than the other models, suggesting it might not be as effective at classifying all classes.

Conclusion

The application of various data analysis techniques, including classification, regression, clustering methods, and Artificial Neural Networks to the data collected from schools in Afghanistan has proven to be particularly effective.

The study successfully identified and analyzed the factors that influence learning and their positive impact on students across different grade levels, school types, and regions. The marks in the 10th and 11th grade are, obviously, the most important features influencing the 12th grade results, but the study time, family situation (such as mother's job and family support of the studies) and the work-life balance (friends, sport and free-time) also are of great importance. The machine learning algorithms allowed to split the students according to their achievements both in two- and several classes that can help educational institutions to predict students' academic success in the future and help those students who fall into «struggling» students.

One of unexpected important factors which impact student's academic success is mother education. If the mother is educated, she influences her kids positively that leads to their academic success. Her profession also influences kids' academic success: if she is a teacher, it is even better for the kids.

The choice of model depends on the specific requirements of the task. For regression tasks, the Random Forest model seems to be the best choice, while for binary classification tasks, all three models perform well, with the SVM model slightly outperforming the others. For multi-class classification tasks, the XGBoost model appears to be the most effective.

In the future it would worth analyzing the future (university) success of the students as well as broadening the geography of the study by adding more data from other Afghan schools.

References

1. Al-Juboori A.M.O. Student performance prediction using deep belief // CoHE Thesis Center. 2019.
2. Altabrawee H., Osama A., Qaisar S. (2019). Predicting students' performance using machine learning techniques // Journal of university of babylon for pure and applied sciences. 2019. № 27. pp. 194-205.
3. Cherry K. How social support contributes to psychological health. Ed. by Morin A. 2023. 4 p. https://www.verywellmind.com/social-support-for-psychological-health-4119970
4. Feng K., Xinyu G., Sajjad S., Kairong Ya., Jingjing Zh. How is emotional intelligence linked to life satisfaction? The mediating role of social support, positive affect and negative affect // Journal of happiness studies. 2019. № 20. pp. 2733-2745.
5. Ge X. Emotion matters for academic success// Education Tech Research and development. 2021. № 69. pp. 67-70
6. Gebauer M.M., McElvany N., Köller O. Cross-cultural differences in academic self-efficacy and its sources across socialization contexts // Social psychology of education. 2021. № 24. pp. 1407-1432.
7. Ghaleb A., El Ref G., Kaba A., Eletter S. The Impact of demographic characteristics on academic performance: face-to-face learning versus distance learning implemented to prevent the spread of COVID-19 // The International review of research in open and distributed learning. 2021. № 22(1). pp. 91-110.
8. How to study effectively: 12 secrets for success. Oxford: Oxford Learning, 2017. https://www.oxfordlearning.com/how-to-study-effectively/
9. Kang X., Matthews S., Yip V. Language and nonlanguage factors in foreign language learning: evidence for the learning condition hypothesis // NPJ Sciens of learning. 2021. № 6. P. 28.
10. Kristensen S.-M., Bogsnes-Larsen T.M., Bjørnøy U.H., Danielsen G.A. Academic stress, academic self-efficacy, and psychological distress: a moderated mediation of within-person effects // Journal of youth and adolescence. 2023. № 52. pp. 1512-1529.
11. Lei H., Cui Yu., Chiu M.M. The relationship between teacher support and students' academic emotions: a meta-analysis // Frontiers in psychology. 2018. Vol. 8. 11 p. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02288/full
12. McCann C. Why you need emotional intelligence to succeed at school // Psychology today. 2020. 4 p. https://www.psychologytoday.com/us/blog/dealing-emotions/202006/why-you-need-emotional-intelligence-succeed-school/

13. Perez M.P., McAllister D.E., Diaz K.G. A longitudinal study of the gender gap in mathematics achievement: evidence from Chile // Educ stud math. 2021. № 107. pp. 583-605.

14. Pojon M. Using machine learning to predict student performance: magister`s diss. Tampere: University of Tampere, 2017. 39 p.

15. Sidole M.W., Olamo T.G., Ayele Z.A. An Investigation on students' attitude towards using mother tongue as a medium of instruction and their awareness about its benefits: Sidaamu Afoo in focus // IOSR journal of humanities and social science. 2019. Vol. 24. Iss. 3. Ser. 4. pp. 61-69

16. Smith G. Factors influencing the speed and success of language learning. 2023. https://mangolanguages.com/resources-articles/what-factors-influence-language-learning/

17. Vijayalakshmi V., Venkatachalapathy K. Comparison of predicting student`s performance using machine learning algorithms // International journal of intelligent systems and applications. № 11.pp. 34-45.

18. Wang M.T., Guo J., Degol J.L. The role of sociocultural factors in student achievement motivation: a cross-cultural review // Adolescent Res Rev. 2020. № 5. pp. 435-450.

19. Wolters C.A., Brady A.C. College students’ time management: a self-regulated learning perspective // Educational psychology review. 2021. № 33. pp. 1319-1351.

20. Wook M., Yahaya Y.H., Wahab N., Isa M.R.M., Awang N.F., Seong H.Y. // Predicting NDUM student's academic performance using data mining techniques, 2009: mat. of the II Inter. conf. on computer and electrical engineering (2009, Dubai). pp. 357-361.

21. Xerri, M.J., Radford, K. & Shacklock, K. Student engagement in academic activities: a social support perspective// High Educ. 2018. № 75. pp. 589-605.

22. Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms // Smart learn. Environ. 2022. № 9. 11 p.

Использование методов машинного обучения для моделирования образовательных данных Афганистана

Азизи Бахрам
Независимый исследователь
Национальный исследовательский университет ИТМО
Санкт-Петербург, Россия
bahramuddinazizi@gmail.com
ORCID 0000-0000-0000-0000

Ольга Борисовна Егорова
Кандидат филологических наук, доцент ВЩЦК
Национальный исследовательский университет ИТМО
Санкт-Петербург, Россия
oegorova@itmo.ru
ORCID 0000-0002-5967-5638

Поступила в редакцию 01.12.2024
Принята 24.01.2025
Опубликована 28.02.2025

УДК 004.428.092.8:371.491.1(945.53)
DOI 10.25726/p0203-5421-9376-e
EDN BTOKYQ

ВАК 5.8.1. Общая педагогика, история педагогики и образования (педагогические науки)
OECD 05.03.HA. EDUCATION & EDUCATIONAL RESEARCH

Аннотация

В данной работе исследуются различные факторы, влияющие на успеваемость школьника на этапе получения среднего образования в Афганистане. Безусловно, учебные достижения играют решающую роль при определении будущей профессии школьника, а также его целей в академической или профессиональной среде. На академическую успешность влияют не только непосредственно учебные показатели, но и другие факторы, включая индивидуальные различия. В этой работе были использованы различные алгоритмы машинного обучения для анализа факторов, влияющих на оценки ученика в последнем классе школы, в частности, обучение с учителем (регрессия и классификация) и без учителя (алгоритмы кластеризации). Цель исследования заключалась в том, чтобы определить, какие факторы существенно влияют на успеваемость учеников для всей рассматриваемой выборки, какие факторы носят более индивидуальный характер, а также степень их влияния. Результаты показали, что различные алгоритмы машинного обучения могут быть успешно использованы для определения факторов, влияющих на академическую успешность студентов. На финальные оценки школьников, несомненно, влияют результаты учебы в предыдущие годы, а также время, затрачиваемое на обучение. Но очень важно отметить, что семейная ситуация и баланс между учебой и активностью в свободное время также влияют на успехи ученика. В частности, уровень образования и работа матери оказывают большое влияние на академические успехи ученика: если мать ученика образована и, особенно, если она работает учителем, то это существенно повышает шансы школьника быть успешным в учебе.

Ключевые слова

моделирование образовательных данных, факторы академической успешности, методы машинного обучения, нейронные сети, данные о среднем образовании, аналитика образовательных данных, интеллектуальный анализ образовательных данных.

Список литературы

1. Al-Juboori A.M.O. Student performance prediction using deep belief // CoHE Thesis Center. 2019.
2. Altabrawee H., Osama A., Qaisar S. (2019). Predicting students’ performance using machine learning techniques // Journal of university of babylon for pure and applied sciences. 2019. № 27. pp. 194-205.
3. Cherry K. How social support contributes to psychological health. Ed. by Morin A. 2023. 4 p. https://www.verywellmind.com/social-support-for-psychological-health-4119970
4. Feng K., Xinyu G., Sajjad S., Kairong Ya., Jingjing Zh. How is emotional intelligence linked to life satisfaction? The mediating role of social support, positive affect and negative affect // Journal of happiness studies. 2019. № 20. pp. 2733-2745.
5. Ge X. Emotion matters for academic success// Education Tech Research and development. 2021. № 69. pp. 67-70
6. Gebauer M.M., McElvany N., Köller O. Cross-cultural differences in academic self-efficacy and its sources across socialization contexts // Social psychology of education. 2021. № 24. pp. 1407-1432.
7. Ghaleb A., El Ref G., Kaba A., Eletter S. The Impact of demographic characteristics on academic performance: face-to-face learning versus distance learning implemented to prevent the spread of COVID-19 // The International review of research in open and distributed learning. 2021. № 22(1). pp. 91-110.
8. How to study effectively: 12 secrets for success. Oxford: Oxford Learning, 2017. https://www.oxfordlearning.com/how-to-study-effectively/
9. Kang X., Matthews S., Yip V. Language and nonlanguage factors in foreign language learning: evidence for the learning condition hypothesis // NPJ Sciens of learning. 2021. № 6. P. 28.

10. Kristensen S.-M., Bogsnes-Larsen T.M., Bjørnøy U.H., Danielsen G.A. Academic stress, academic self-efficacy, and psychological distress: a moderated mediation of within-person effects // Journal of youth and adolescence. 2023. № 52. pp. 1512-1529.

11. Lei H., Cui Yu., Chiu M.M. The relationship between teacher support and students' academic emotions: a meta-analysis // Frontiers in psychology. 2018. Vol. 8. 11 p. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02288/full

12. McCann C. Why you need emotional intelligence to succeed at school // Psychology today. 2020. 4 p. https://www.psychologytoday.com/us/blog/dealing-emotions/202006/why-you-need-emotional-intelligence-succeed-school/

13. Perez M.P., McAllister D.E., Diaz K.G. A longitudinal study of the gender gap in mathematics achievement: evidence from Chile // Educ stud math. 2021. № 107. pp. 583-605.

14. Pojon M. Using machine learning to predict student performance: magister`s diss. Tampere: University of Tampere, 2017. 39 p.

15. Sidole M.W., Olamo T.G., Ayele Z.A. An Investigation on students' attitude towards using mother tongue as a medium of instruction and their awareness about its benefits: Sidaamu Afoo in focus // IOSR journal of humanities and social science. 2019. Vol. 24. Iss. 3. Ser. 4. pp. 61-69

16. Smith G. Factors influencing the speed and success of language learning. 2023. https://mangolanguages.com/resources-articles/what-factors-influence-language-learning/

17. Vijayalakshmi V., Venkatachalapathy K. Comparison of predicting student`s performance using machine learning algorithms // International journal of intelligent systems and applications. № 11.pp. 34-45.

18. Wang M.T., Guo J., Degol J.L. The role of sociocultural factors in student achievement motivation: a cross-cultural review // Adolescent Res Rev. 2020. № 5. pp. 435-450.

19. Wolters C.A., Brady A.C. College students' time management: a self-regulated learning perspective // Educational psychology review. 2021. № 33. pp. 1319-1351.

20. Wook M., Yahaya Y.H., Wahab N., Isa M.R.M., Awang N.F., Seong H.Y. // Predicting NDUM student's academic performance using data mining techniques, 2009: mat. of the II Inter. conf. on computer and electrical engineering (2009, Dubai). pp. 357-361.

21. Xerri, M.J., Radford, K. & Shacklock, K. Student engagement in academic activities: a social support perspective// High Educ. 2018. № 75. pp. 589-605.

22. Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms // Smart learn. Environ. 2022. № 9. 11 p.