Low-achievement risk assessment with machine learning

Andrea Zanellati[1,*], Stefano P. Zingaro[1] and Maurizio Gabbrielli[1]

[1]Università di Bolonga, via Zamboni 33, Bologna, 40126, Italy

Abstract

In this work, we propose a method for assessing the risk of low-achievement in secondary school with data collected from the Italian ministry of education. Low-achievement is a phenomenon whereby a student, despite completing his or her education, does not reach the level of competence expected by the school system. We train three machine learning models on a large, real dataset through the INVALSI large-scale assessment tests and compare the results in terms of predictive and descriptive performance. We exploit data collected in end-of-primary school mathematics tests to predict the risk of low-achievement at the end of compulsory schooling (5 years later). The promising results of our approach suggest that it is possible to generalise the methodology for other school systems and for different teaching subjects.

Keywords

low-achievement, performance prediction, assessment test, machine learning

1. Introduction

Low-achievement at school is a widespread phenomenon which has long-term consequences, both for the individual and for society as a whole. In 2016, above 28% of students across Organization for Economic Co-operation and Development (OECD) countries underscored the minimum level of proficiency in at least one of the three core subjects according to the Programme for International Student Assessment (PISA), which are English reading and comprehension, mathematics, and science [1]. Low-achievement is strongly related to school dropout, i.e., the discontinuation of education [2], and impact on the cultural and professional growth of the individual and citizen [3, 4]. Indeed, school performance in first grade is already a significant indicator of future high dropout risk. In 2019, a study conducted by the National Institute for Assessment of the Education System (INVALSI) found that 20% percent of Italian students had a lower-than-expected achievement and, eventually, dropped out of school [5]. This way, despite the exterior appearance and, in some cases, despite the sufficient marks, the students do not reach the adequate level of knowledge which later on will be needed to successfully continue the studies or to start a professional career. In this perspective, low achievement can be considered an "implicit" form of school dropout: although some students do not occur into explicit early leaving from school, the result is a lack of school effect on their skills acquisition.

An important indicator for school dropout is the ELET rate, which measures the percentage of Early Leavers from Education and Training [6]. It measures a severe condition of educational exclusion which refers to young people between 18 and 24 with a qualification lower than upper secondary. Early leavers from education and training are more likely to be unemployed or employed in low-paid jobs with few or no prospects for training and further career progression; they are more prone to social exclusion and to experience lower levels of health, well-being and life satisfaction; they are also more likely to experience limited civic participation. In [7] the European Commission indicated as one of the main targets to be achieved in education the reduction of the ELET rate from 15% to 10% in the decade 2010–2020. While this target was reached in several European countries, Italy in 2019 still had an ELET rate of 13.5%[8].

To counteract dropout as soon as possible and to detect low-achievement, we address the following research questions:

> **RQ1** Is it possible to quantitatively represent a student's knowledge level and build a model of his or her skill attainment?
>
> **RQ2** Is it possible to develop a suitable AI-tool to predict, at an early stage, the risk of low-achievement at secondary school for primary school students?

In the following, we present a case study, focusing on the Italian context and using data collected from the INVALSI national large-assessment tests in mathematics. In particular, from these tests, we aim to extract the relevant features related to students' learning in terms of their skill and competence level performance.

In **RQ2**, we refer to "early stage" meaning to detect risk as soon as possible, i.e., several years in advance, so

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.
andrea.zanellati2@unibo.it (A. Zanellati); stefano.zingaro@unibo.it (S. P. Zingaro); maurizio.zingaro@unibo.it (M. Gabbrielli)
0000-0001-6171-0397 (A. Zanellati); 0000-0002-8462-5651 (S. P. Zingaro); 0000-0003-0609-8662 (M. Gabbrielli)

that appropriate countermeasures can be taken, and to design an intervention aimed at reducing risk when it is detected. Concretely, we develop three models able to predict the risk of low-achievement at K-10, using student data at K-5.

In selecting the models, we strive for a balance between descriptive and predictive performance. Indeed, we want our solution to be both interpretable and robust. Hence, we consider state-of-the-art machine learning techniques that proved to be effective in preliminary experiments [9]: random forests and neural networks [10, 11]. On the one hand, we exploit random forests to extract rules that facilitate the process of interpreting the outcomes of the research and, on the other hand, we test neural networks for flexibility, e.g., exploring non-linear correlations, and performance gain.

2. Related work

Low-achievement is a widely studied phenomenon in the social sciences and education [12, 13]. The problem has also been addressed in terms of predictive models for low-achievement—or dropout risk—for both high school and college students. These models exploit different machine learning techniques, including supervised learning, e.g., random forests, support vector machine and Bayesian network, unsupervised learning, e.g., k-means and hierarchical clustering, and recommender systems, e.g., collaborative filtering [14, 15]. Moreover, several kinds of data have been used to tackle the problem. In [16] the dataset for building the predictive model uses demographic data of the students and their grades. Other studies are based on students performance, i.e., grades, collected during first semester courses [17, 18]. Some include behavioural data supplemented with other features related to learning results [19], in a mix of cognitive and non-cognitive characteristics. In some studies data collected through large-scale assessment tests were used to design predictive models of student performance through several machine learning techniques. In [20], for example, the authors refer to data collected through the PISA international large-scale assessment tests.

In this scenario, we aim to contribute to the research field of AI-based education solutions by presenting a case study for predicting the risk of low-achievement of high school students using their performance data collected during primary school. As a minor contribution, we extract features directly related to students' learning in terms of knowledge and skills, privileged indicators for the study of learning [21], thus proposing a knowledge-based method for encoding students' learning. We believe that this element can improve the interpretability of the results and make this tool useful for students, teachers and instructional coordinators.

3. Methodology

The INVALSI dataset is the result of a large-scale assessment administered in Italy since the school year 2002/03 at the levels K-2, K-5, K-8, K-10, and K-13. Students are tested in Italian, Maths and English. In addition to this, there is a student-survey aimed at gathering information on the social, economic and cultural context. A strength point of the dataset is represented by the possibility to longitudinally link data collected in different school years [22] starting from 2011. In our case study, we considered data on maths test from two cohorts of students: K-5 of the 2012/13 school year and K-5 of the 2013/14 school year. For the same students, we collected data from five years later at grade K-10, to be used for the definition of the low-achievement target, i.e., the students grade in the test is less than or equal to 2 on a scale from 1 to 5. After merging K-5 datasets with their correspondent K-10 targets, the K-5 2012/13 cohort is made up of 351746 students, while the K-5 2013/14 cohort of 354987 students.

There are several features in the dataset and we applied a feature selection process to determine a subset of relevant features. The datasets also contain a boolean feature for each test item, where the students' answers correctness are recorded. To enable the use of our predictive models on different cohorts of students and to provide a coherent representation of their learning in terms of areas of knowledge and skills, it is necessary to release the dataset from the individual items that constitute a certain test. Therefore, we used a knowledge-based approach considering the items classification in terms of areas, processes and macro-processes according to the INVALSI framework for the design of math tests.

In Table 1, we give for reference an overview of the areas, processes, and macro-processes that have been used in the encoding of the questions.

We define one new variable for each area, process, and macro-process. Each of these new features takes the value corresponding to the percentage of correct answers provided by the student for that specific group of items, namely, correctness rate. Last, we concatenate the computed values to obtain a new flattened representation of learning, where each item is a possible indicator and not its unique representative. Following our strategy, we represent each student's learning in the space of fifteen (15) dimensions, as shown in Table 2.

We use two techniques to develop our AI-based tool. The first one is Random forest (RF) [10], which is widely used in Educational Data Mining for the high degree of explainability and effortless interpretation of the results. We trained our models through bootstrap aggregating (bagging) to reduce the overfitting of dataset and increase precision. To tune the model, we performed a grid search.

The second technique is based on neural networks,

Table 1
Maths INVALSI framework for question encoding.

Areas
(NU) Numbers (SF) Space and figures (DF) Data and forecasts (RF) Relations and functions
Process
(P1) Know and master the specific contents of mathematics (P2) Know and use algorithms and procedures (P3) Know different forms of representation and move from one to the other (P4) Solve problems using strategies in different fields (P5) Recognize the measurable nature of objects and phenomena in different contexts and measure quantities (P6) Progressively acquire typical forms of mathematical thought (P7) Use tools, models and representations in quantitative treatment information in the scientific, technological, economic and social fields (P8) Recognize shapes in space and use them for problem solving
Macro-process
(MP1) Formulating (MP2) Interpreting (MP3) Employing

Table 2
Example of the student's learning final encoding.

Id	NU	SF	DF	RF	P1	P2	P3	P4	P5	P6	P7	P8	MP1	MP2	MP3
1	0.86	0.75	0.90	0.80	0.71	0.80	1.00	0.89	1.00	0.67	0.91	0.75	0.81	0.73	0.94
2	0.50	0.25	0.50	0.53	0.29	0.60	0.50	0.22	1.00	0.33	0.73	0.25	0.50	0.47	0.44

which has recently become widespread also in the field of Educational Data Mining and has also been applied in predictive models for student performance [23]. To deal with our hybrid dataset, we firstly include a preprocessing step, aimed at encoding the values of categorical variables into numerical values with a "one-hot" encoding algorithm. After preprocessing, we implemented two neural networks based on different data transformation approaches. Categorical Embeddings (CE), is a neural network that treats the input depending on its type: categorical inputs are passed through an embedding layer, numerical ones are fed to a dense layer. Feature Tokenizer Transformer (FTT) [11] is able to identify the input or the group of inputs that most influence the output, thanks to attention maps. It is a more complex architecture that exploit a feature tokenizer function to extract tokens from the input and then fed these tokens to a Transformer architecture [24] for classification.

4. Experimental results

We carried out all the experiments using the Google Colaboratory Notebook environment, with the Python programming language and popular machine learning libraries, such as scikit-learn and pandas.

The dataset for all the experiments was pre-processed cleaning features with many missing values, highly correlation (computed by R^2 measure above 0.5) or specifically referred to a cohort of students, preventing the model to be transferred to new cohorts (e.g., identification code for a class). This features selection process, together with the engineering of the features related to the items in the tests, results in a set of 34 features, which refers both to socio-economic and cultural context, demographic data and learning dimension. For the definition of the training set we used the data from 2012/13 K-5 cohort. For the models based on neural networks we split this cohort to generate both training and validation sets (split in 80% and 20% respectively). Finally, we used the K-5 2013/14 cohort to test and measure the model performance. This allowed us to evaluate the validity of the proposed learning encoding that is, the effectiveness of abstraction from specific items to learning in terms of areas, processes and macro-processes. The dataset is unbalanced between underachievement/non-underachievement classes; therefore balancing techniques were applied. In the development of the RF models, a random undersampling technique was used, implemented in the `imblearn` library. We trained neural networks using a weighted random sampler, that samples the data to balance classes ratio in the training batches.

Predictive models performance

In Table 3, we present the overall results on the test dataset of the above mentioned models: RF, CE, and FTT. For RF, we considered the best hyper-parameters setting

Table 3
Predictive performance of the proposed models on the external validation test set (year 2015/16).

Models	Accuracy	Precision	Recall
Random Forest	0.77	0.62	0.67
CE network	0.76	0.76	0.76
FTT network	0.78	0.77	0.78

determined with the grid search technique: 50 estimators in the forest, trained with 30% of random samples, 60% of random features and max depth set to 11. The FTT outperforms the other predictive models with accuracy, precision and recall between 77% and 78%.

5. Conclusion

Our results suggest that the challenge of predicting low-achievement risk for primary and secondary school students can be effectively addressed through the use of well-curated datasets and the choice of reliable predictive models. Our abstract representation of (INVALSI) tests and the related encoding for the student achievement allowed us to transfer the trained models on different cohorts and therefore to obtain a accurate prediction. We believe that the ability to predict low school achievement with reasonable accuracy five years in advance offers a practical tool for policy makers, managers and educators.

We are interested in extending our work in several directions. First, we want to verify the transferability of the proposed methodology to other disciplines, using a representation for students' learning similar to the one proposed in this work. Second, we want to increase the quality of the information provided as input to the predictive models, e.g., by collecting more data and by integrating new data sources. We aim to improve the learning encoding—thus the feature extraction process—in a way that is not knowledge-based to limit the bias. Finally, we want to deepen the interpretability of the results of our models, by analysing the feature importance computed on RF model and comparing it with the interpretation of the weights that define the neural networks we have used.

References

[1] OECD, Who and Where are the Low-Performing Students?, OECD Publishing, 2016.
[2] D. D. Curtis, J. McMillan, School non-completers: Profiles and initial destinations, 2008.
[3] K. L. Alexander, D. R. Entwisle, L. S. Olson, Schools, achievement, and inequality: A seasonal perspective, Educational evaluation and policy analysis 23 (2001) 171–191.
[4] S. J. Ingels, T. R. Curtin, P. Kaufman, M. N. Alt, X. Chen, et al., Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later., Eric, 2002.
[5] R. Ricci, La dispersione scolastica implicita, 2019.
[6] S. Flisi, V. Goglio, E. C. Meroni, E. Vera-Toscano, School-to-work transition of young individuals: what can the elet and neet indicators tell us, Luxembourg: Publications Office of the European Union, EUR-Scientific and Technical Research Reports (2015).
[7] E. Commision, Europe 2020 a strategy for smart, sustainable and inclusive growth, 2010.
[8] Istat, Livelli di istruzione e occupazione nazionali, 2020.
[9] A. Zanellati, S. P. Zingaro, M. Gabbrielli, Student low achievement prediction, in: M. M. Rodrigo, N. Matsuda, A. I. Cristea, V. Dimitrova (Eds.), Artificial Intelligence in Education, Springer International Publishing, Cham, 2022, pp. 737–742.
[10] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.
[11] Y. Gorishniy, I. Rubachev, V. Khrulkov, A. Babenko, Revisiting deep learning models for tabular data, 2021. arXiv:2106.11959.
[12] R. Cassen, G. Kingdon, Tackling low educational achievement, Joseph Rowntree Foundation, 2007.
[13] D. C. Geary, Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics, Journal of developmental and behavioral pediatrics: JDBP 32 (2011) 250.
[14] J. L. Rastrollo-Guerrero, J. A. Gomez-Pulido, A. Durán-Domínguez, Analyzing and predicting students' performance by means of machine learning: A review, Applied sciences 10 (2020) 1042.
[15] B. Albreiki, N. Zaki, H. Alashwal, A systematic literature review of student performance predic-

tion using machine learning techniques, Education Sciences 11 (2021) 552.

[16] S. Kotsiantis, C. Pierrakeas, P. Pintelas, Predicting students'performance in distance learning using machine learning techniques, Applied Artificial Intelligence 18 (2004) 411–426.

[17] M. A. Al-Barrak, M. Al-Razgan, Predicting students final gpa using decision trees: a case study, International journal of information and education technology 6 (2016) 528.

[18] Z. Ibrahim, D. Rusli, Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007.

[19] S. Sultana, S. Khan, M. A. Abbas, Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts, International Journal of Electrical Engineering Education 54 (2017) 105–118.

[20] A. Pejić, P. S. Molcer, K. Gulači, Math proficiency prediction in computer-based international large-scale assessments using a multi-class machine learning model, in: 2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY), IEEE, 2021, pp. 49–54.

[21] L. K. Baartman, E. De Bruijn, Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence, Educational Research Review 6 (2011) 125–134.

[22] L. Branchetti, F. Ferretti, A. Lemmo, A. Maffia, F. Martignone, et al., A longitudinal analysis of the italian national standardized mathematics tests, in: CERME 9-Ninth Congress of the European Society for Research in Mathematics Education, 2015, pp. 1695–1701.

[23] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, B. Navarro-Colorado, A systematic review of deep learning approaches to educational data mining, Complexity (2019).

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, et al., Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.