Interpretable Deep Learning for University Dropout Prediction

Máté Baranyi
Budapest University of Technology and Economics
baranyim@math.bme.hu

Marcell Nagy
Budapest University of Technology and Economics
marcessz@math.bme.hu

Roland Molontay
MTA-BME Stochastics Research Group
molontay@math.bme.hu

ABSTRACT

The early identification of college students at risk of dropout is of great interest and importance all over the world, since the early leaving of higher education is associated with considerable personal and social costs. In Hungary, especially in STEM undergraduate programs, the dropout rate is particularly high, much higher than the EU average. In this work, using advanced machine learning models such as deep neural networks and gradient boosted trees, we aim to predict the final academic performance of students at the Budapest University of Technology and Economics. The dropout prediction is based on the data that are available at the time of enrollment. In addition to the predictions, we also interpret our machine learning models with the help of state-of-the-art interpretable machine learning techniques such as permutation importance and SHAP values. The accuracy and AUC of the best-performing deep learning model are 72.4% and 0.771, respectively that slightly outperforms XGBoost, the cutting-edge benchmark model for tabular data.

CCS CONCEPTS

• **Applied computing → Education**; • **Information systems → Decision support systems**.

KEYWORDS

dropout prediction; higher education; neural networks; interpretable machine learning; explainable artificial intelligence

ACM Reference Format:
Máté Baranyi, Marcell Nagy, and Roland Molontay. 2020. Interpretable Deep Learning for University Dropout Prediction. In *The 21st Annual Conference on Information Technology Education (SIGITE '20), October 7–9, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3368308.3415382

1 INTRODUCTION

Student dropout and delayed completion are some of the most burning issues in higher education all over the world. Approximately every third student drops out from higher education worldwide [14, 15, 38], which is associated with considerable personal and social costs [21, 31]. In Hungary, the graduation rate is particularly low, one of the lowest among OECD and EU countries [12, 38].

Understanding and identifying the factors affecting students' university performance, and detecting at-risk students are in the focus of educational research for decades [1, 5, 11, 23, 32]. Moreover, due to the rise of data-driven approaches and the vast amount of data stored in educational administrative systems, a tremendous number of predictive analytical educational research papers and artificial intelligence-based decision support systems have been published in the last few years, for systematic reviews, we refer to [10, 13, 19, 33].

The majority of the related works use traditional machine learning models (e.g. logistic regression, k-nearest neighbors, and decision tree-based ensemble models) for dropout prediction [6, 7, 9, 34]. Deep neural networks (DNN) are not frequently applied on tabular data, due to the fact that such datasets are usually not large and complex enough for these models to achieve high performance, moreover, DNNs are more computationally expensive and they require more fine-tuning and expertise, besides these models are less interpretable than the traditional ones. However, there is a growing number of deep learning models that are designed for tabular data (e.g. TabNet [4], RLNs [36], and NODE [29]), hence many researchers started to incorporate these models in their analyses. Also in educational data science some papers already successfully applied deep neural networks for student performance and dropout prediction. Xing and Du constructed a dropout prediction model for identifying at-risk students in MOOCs [39]. Kim *et al.* proposed a deep learning-based model, called GritNet, that handles student performance prediction as a sequential event prediction problem [16]. A probabilistic neural network was introduced by Mason *et al.* to predict engineering student retention [24]. Recently, Agrusti *et al.* trained a convolutional neural network to predict student dropout [2]. Various other network architectures (e.g. denoising autoencoders) have been used to predict student retention and identify at-risk students [3, 20, 28], nevertheless a common conclusion can be drawn: deep learning models frequently outperform traditional machine learning models for student dropout prediction.

The present study joins the series of works aiming to build accurate dropout prediction models at the Budapest University of Technology and Economics. The last two authors of the present paper have used traditional machine learning classifiers to predict students' final academic performance based on pre-enrollment achievement measures [26], later a decision support tool was built on the results, which not only returns the predicted probability of dropping out but also shows the key factors affecting the individual predictions [27]. Moreover, in a follow-up paper, Kiss *et al.* investigated how the prediction can be improved using first-semester

SIGITE '20, October 7–9, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7045-5/20/10…$15.00
https://doi.org/10.1145/3368308.3415382

performance indicators, i.e. studied the incremental predictive validity of first-semester university performance indicators on retention over pre-enrollment achievement measures [17].

This work extends the aforecited papers in several directions. First of all, the main focus of this work is to investigate, whether deep neural networks with proper hyperparameter optimization are able to identify students at risk of dropout. To this end, we compare the performance of DNNs to state-of-the-art tree-based machine learning models such as XGBoost. Another key contribution of this work is that we do not only give a prediction on the final academic status, but also use machine learning interpretability and explainability tools such as SHAP values [22], and permutation importance. These tools help us understand and interpret the decisions of the models. Furthermore, they also provide global and local explanations of the predictions that can provide insights for the decision-makers of higher education. Moreover, by highlighting the factors that decrease the estimated probability of graduation, it can also help students identify the skills they need to master in order to be successful in higher education.

The main objectives of this work can be summarized as follows:

(1) Using deep neural networks to predict the students' final academic performance (graduation or dropout) as accurately as possible, and compare their performance to other machine learning models such as XGBoost and random forest.
(2) Using interpretable machine learning and explainable artificial intelligence tools to understand the decisions of the models, moreover to find what the key factors are that affect academic success.

After providing some background information on Hungarian higher education in Section 2, the rest of the paper follows the workflow depicted in Figure 1. Namely, Section 3 is devoted to data description, the machine learning models together with their evaluation can be found in Section 4, while Section 5 is devoted to model interpretation. Section 6 concludes the work.

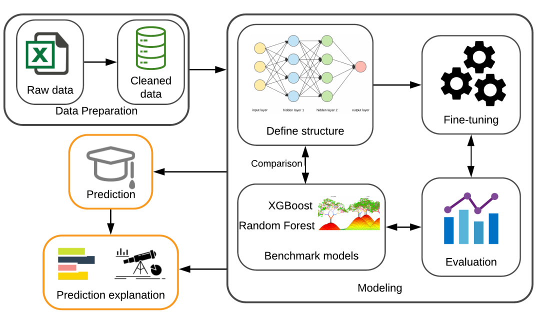

Figure 1: The workflow of this project.

2 BACKGROUND ON HUNGARIAN HIGHER EDUCATION

To make the paper easier to follow, here we briefly present the most important aspects of the transition to higher education in Hungary. A more detailed description of the Hungarian higher education admission procedure can be found in [26].

At the end of secondary education, students take an exit exam, called *matura*, that is a matriculation exam as well, meaning that it serves as a university entrance exam at the same time, providing that the student applied to higher education. Matura consists of (at least) five separate sub-exams of core subjects: mathematics, history, Hungarian language and literature, a foreign language, and (at least) one chosen subject. Any sub-exam can be taken at a normal or advanced level depending on the student's choice.

Students applying to higher education, collect a so-called University Entrance Score (UES), that is a composition of *Matura points (MP)*: scores of two university program-specific matura sub-exams, required by the university program; *Study points (SP)*: grade point average in high school together with the results of all the five matura sub-exams; and some additional *Extra points (EP)* (e.g. reward for advanced level exam and foreign language certificate). Each undergraduate program determines a Minimal University Entrance Score (MUES), and only those students can gain admission whose UES is greater than the minimum score.

There are two possible ways to calculate the UES: the *general* and the *specialist* calculations, and the final entrance score will be the maximum of these two scores. The general calculation ($UES = MP + SP + EP$) rather measures the general knowledge of the students, because it takes into account the matura scores and the grades of all the five core subjects. On the other hand, the specialist calculation ($UES = 2 \cdot MP + EP$) rather focuses on the program-specific knowledge of the students, e.g. in case of an engineering student the specialist knowledge involves mathematics and a natural science subject, typically physics, hence the specifically calculated score is independent of the student's scores and grades in humanities such as history and Hungarian.

Figure 2 shows that students who focus only on the program-specific subjects are less likely to graduate than those who have great general knowledge and who are open-minded about subjects that are not program-specific. It suggests that *generalist* students may have learning skills that are advantageous in higher education. The figure also illustrates that the majority of female students are *generalists* and that males graduate less likely than their female peers.

3 DATA DESCRIPTION

This study is based on the data of undergraduate students at the Budapest University of Technology and Economics (BME), the flagship of Hungarian engineering higher education. The filtered and cleaned data consist of 8,319 students who enrolled between 2013 and 2019 to an undergraduate program at BME and finished their studies either by graduation or dropping out. It means that there are some students who enrolled since 2017 but dropped out in their first year, on the other hand graduated students could not have enrolled after 2016, since programs are at least six semesters long. Due to an upgrade of the administration system, there were some incomplete rows in the raw data, we imputed the missing fields using multiple imputation.

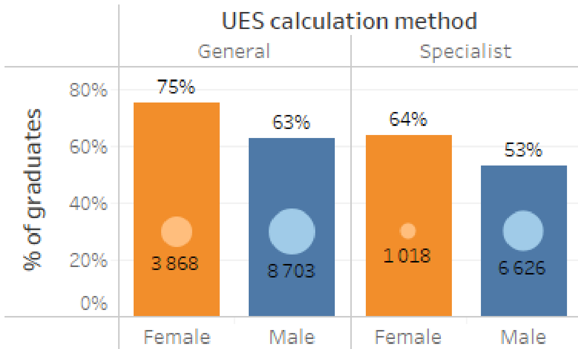

Figure 2: The percentage of male and female graduates, enrolled between 2010 and 2015, broken down by the UES calculation method. The corresponding number of students for each group are shown inside the bars by the circles together with the numbers.

The features of the data set are listed in Table 1. We only used features available at the time of enrollment, while most of the related works solve a less difficult problem by using first-semester achievement measures as well. While most of the variables are straightforward, there are a few that require some explanation. The variable *Re-enrolled* indicates whether the student has already been enrolled in the university before, dropped out, and then enrolled again. Unfortunately, due to some recent policy changes, the number of re-enrolling students increases from year to year. Moreover, the *Surplus entrance score* attribute measures how much the given student outperformed the minimum score (MUES); that is defined as the difference between the UES and MUES weighted by the MUES, since exceeding a high minimum score by a unit reflects a better performance than exceeding a low minimum score by the same unit. The so-called *Years elapsed* variable is defined as the years elapsed between the time of the matura examination and the university enrollment. This variable obviously highly correlates with the *Re-enrolled* variable. We also introduced a new variable that encodes the location type of the high school of the students (capital city, city with county rights, town). We also defined a score for the foreign language certificates that takes into account the number and the level of language certificates.

Table 1: Summary of features.

Feature Class	Feature name	Type
University program related	Student ID	Nominal
	Program ID	Nominal
	Faculty of program	Categorical
	State-funded (or fee-paying)	Binary
	Re-enrolled	Binary
Target variable	Final status	Binary
High school performance related	University entrance score	Real
	Surplus entrance score	Real
	Matura points	Real
	Study points	Real
	Extra points	Real
	Competition achievement	Binary
	UES calculating method	Binary
Matura exam results	Hungarian Language and Literature	Real
	Mathematics	Real
	History	Real
	Foreign Language	Real
Average of grades	Mathematics	Real
	Hungarian Language	Real
	Hungarian Literature	Real
	History	Real
	Foreign Language	Real
	Chosen science subject	Real
Foreign language certificate	Score (based on the number and level of certificates)	Real
Personal details	Gender	Binary
	Location type of high school	Nominal
	Years elapsed between the time of matura and enrollment	Real

4 MODELING AND EVALUATION

We aim to predict whether a student graduates or drops out based on their pre-enrollment achievement measures, thus the problem at hand is a supervised learning problem, more precisely, a binary classification problem. The inputs are the attributes of incoming students that are known at the time of enrollment (see Table 1), while the output is the final academic performance measured as a binary variable (1: graduated, 0: dropped out). The class distribution is relatively balanced in the dataset. The reason for having a balanced dataset is that we have data of students who enrolled between 2013 and 2019, which means that students who enrolled since 2017 could not graduate yet, but they might have dropped out already. This can be considered as a natural oversampling technique.

4.1 Modeling

We apply several machine learning models to provide predictions on the final academic status (graduated/dropped out): for benchmark models, we use random forest and XGBoost, furthermore, we train three neural network models.

The first neural network model is a simple fully connected deep neural network (FCNN), which is able to achieve relatively high performance on tabular datasets as the cited works present [3, 28, 39]. A recent paper [4] proposes a novel interpretable canonical deep tabular data learning architecture, called TabNet, and the authors claim that it can even outperform decision tree based machine learning models on a wide range of tabular data, thus our second deep neural network model utilizes TabNet architecture. Our third model, referred to as BaggingFCNN, is a bagging ensemble metamodel that aggregates 10 (individually trained) copies of the first FCNN model.

4.2 Optimization and evaluation

We conducted a hyperparameter optimization with the *hyperas* Python package [30] for the FCNN model. The tested hyperparameters of our FCNN model are shown in Table 2, where the finally chosen parameters are highlighted in **bold**. The hyperparameter optimization was evaluated on the validation set (749 instances) and the final performance evaluation of the models was carried out on the test set (832 instances).

Table 2: The tested hyperparameter space for the FCNN.

hyperparameter	Values
Depth[1] of the NN	{1, 2, 3, 4, 5, 6, **7**, 8, 9}
Size of the FC layers	{35, **40**, 45, 50, 55}
Activation functions	{ReLU, Sigmoid, **SeLU**}
Additional layer between FC layers	DropOut[2], **BatchNormalization**, AlphaDropOut
Kernel initializer	**Glorot normal**[3], Glorot uniform[4], constant ones
Optimizer	SGD, Adam, **RMSprop** with different learning rates
Batch size	{**16**, 32, 64, 128 }

The *hyperas* package uses the Tree-structured Parzen Estimator (TPE) hyperparameter optimization algorithm. TPE iteratively modifies the sampling distribution of the hyperparameter space, preferring those set-ups that worked better earlier. The parameter settings of the algorithm were as follows: the number of iterations was set to 300, each with 60 epochs, the optimized metric was chosen to be the accuracy on the validation dataset. With the above setup, the optimization took around three hours on an i7-5600U CPU running at 2.60GHz but this is heavily subject to the chosen number of iterations, the size of the training data, and the number of trainable model parameters. The validation accuracy of the models against the tuned hyperparameters can be seen in Figure 3. The fact that it is challenging to train a neural network on this dataset, and a lot of training trials end up with the accuracy of a coin-toss indicates that in this paper we solve a non-trivial problem. During

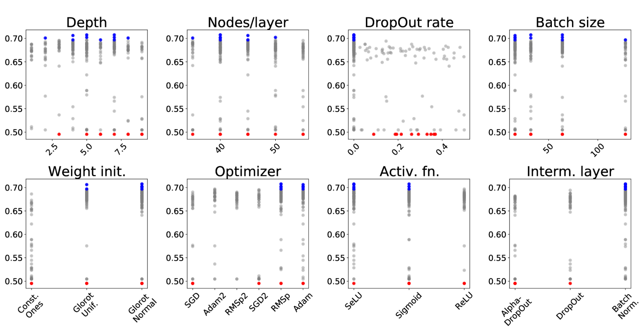

Figure 3: Accuracy on the *validation* set against different hyperparameters. Blue points indicate the best ten, red points indicate the worst ten models.

the hyperparameter optimization, we also tested a few novel techniques, e.g. the AlphaDropout layer and SeLu (scaled exponential linear units) activation function. These techniques are the building blocks of self-normalizing neural networks, which was proposed by Klambauer *et al.* [18] who argue that the use of these methods can enhance the performance of feed-forward neural networks.

The performance of the five chosen models on the test dataset is reported in Table 3, which shows very similar results for all tested models, however, our fully connected deep neural network model slightly outperforms the other models. The receiver operating characteristic curves (ROC) of the models are shown in Figure 4. The area under the curve (AUC) of the models are also similar, being the highest for the FCNN and XGBoost models.

The performance of our models is in line with the ones reported in related works; for example, Plagge achieved an accuracy of 75.7% with their best neural network model for student retention [28], while Alkhasawneh and Hobson could classify 70.1% of the students correctly [3].

Table 3: Different scores on the test dataset.

Model	Accuracy	Precision	Recall	AUC
FCNN	**0.724**	**0.747**	**0.667**	**0.771**
TabNet	0.715	0.742	0.650	0.766
XGBoost	0.707	0.740	0.629	**0.771**
RandForest	0.673	0.723	0.551	0.763
BaggingFCNN	0.719	0.743	0.660	0.770

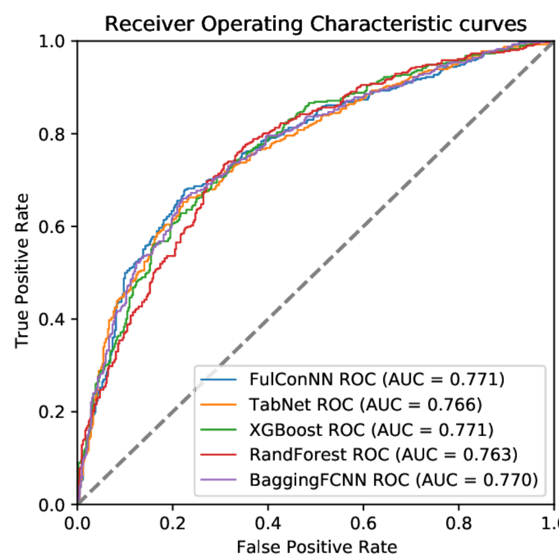

Figure 4: ROC curves of the trained final models, the bagged estimator version, a random forest, and an XGBoost model.

5 MODEL INTERPRETATION

Some machine learning models, especially deep neural networks, are extremely difficult or might even impossible to interpret, that is why some segments of the *industry* are afraid of deploying neural networks based systems, for example, economists prefer linear

[1]Number of fully-connected layers between the input and output layers
[2]The dropout layers were tested with different dropout rates as well
[3]Zero centered truncated normal distribution with standard deviation $std = \sqrt{2/(\#nodes_{in} + \#nodes_{out})}$
[4]Uniform distribution with limits $\pm\sqrt{6/(\#nodes_{in} + \#nodes_{out})}$

regression models for the opposite reason. On the other hand, interpretable machine learning is a new research area of great interest [25]. Here we demonstrate how we can explain the decisions of sophisticated models with explainable artificial intelligence techniques.

We examine a few model interpretability tools to gain a better understanding of the trained model. In the boxplots of Figure 5, we can see how the predicted probability of the graduation changes along the UES in the training set. This is in line with the expected behavior of the model, namely the higher the university entrance score is, the higher the probability of graduation is, which makes the UES a valid predictor of later university performance.

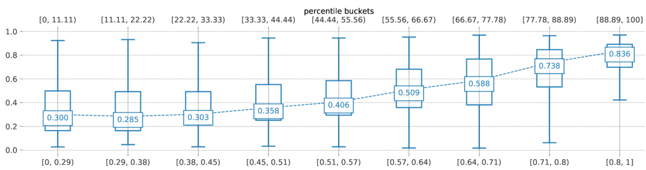

Figure 5: Boxplots of the predicted confidence score of graduation (y-axis) along binned UES groups (x-axis). The UES score is scaled to the $[0, 1]$ interval, and the size of the bins is balanced.

5.1 Permutation importance

Feature importance can be measured by looking at how much the performance of a trained model decreases when the values of a feature are permuted (shuffled), ceteris paribus [8]. Namely, first we train a model and calculate its performance, then we shuffle the values of one feature, and without re-training the model, we again calculate its performance on the shuffled data and investigate how much the performance worsened. Clearly, if a variable is important, then the accuracy of the model will drastically decrease after the permutation. On the other hand, if we shuffle the values of a redundant variable, then the performance of the model will remain the same. The SHAP global feature importance (mean absolute Shapley values across the data) of the 12 most important features are shown in Table 4 together with their permutation importance.

The most important features, according to the permutation importance, are the *Years elapsed*, i.e. the number of years that elapsed between high school graduation and enrollment, the *State-funded* indicator: whether the program is state-funded or fee-paying, the university entrance score, the results in mathematics, and the gender. A few program codes also seem to be important, that is because graduation rates vary across programs.

5.2 SHAP value

We also aim to understand why the machine learning models predict dropout or graduation for an individual student. To this end, similarly to [27], we use the SHapley Additive exPlanations (SHAP) approach [22]. The SHAP value is a state-of-the-art machine learning model explainability tool, that is based on the Shapley value, a cooperative game theoretical concept that quantifies the contribution of each player in a coalition [35].

Table 4: The 12 most important features according to the SHAP global importance and their permutation importances. The ranks are shown in the brackets.

Name	SHAP	Permutation
Years elapsed	0.0473	0.0397 (1)
University entrance score	0.0361	0.0198 (3)
Mathematics avg. grade	0.0349	0.009 (17)
Foreign language	0.0343	0.0104 (13)
Matura points	0.0339	0.0131 (8)
History avg. grade	0.0324	0.0103 (14)
Surplus entrance score	0.0319	0.0099 (15)
Mathematics	0.0315	0.0134 (5)
Program 10	0.0295	0.0153 (4)
State-funded	0.0287	0.0322 (2)
Male	0.0277	0.0132 (7)
Specialist calculation	0.0267	0.0113 (11)

In machine learning settings, the players are the explanatory variables and the v value function of a coalition corresponding to the given subset of variables is defined as follows [37]. Let f denote the machine learning model that predicts the target variable (in our case the final academic performance) for an instance x (in our case an instance is a d-dimensional vector containing the attributes of a student). Let D be the index set of features, i.e. $D = \{1, 2, \dots, d\}$ and S be a subset of D. Then we define $f_S(x)$ as the conditional expectation of $f(x)$ given the values of the X_i attributes belonging to the set S, formally:

$$f_S(x) = \mathbb{E}(f(x) \mid X_i = x_i, \forall i \in S)$$

Note that if S is an empty set, then $f_S(x)$ is the expectation of $f(x)$, i.e. $f_{\{\}}(x) = \mathbb{E}(f(x))$. The expected values are estimated from the data. Using these notations, the value of a contribution of a subset of features is defined as:

$$v(S) = f_S(x) - f_{\{\}}(x),$$

which is the change in prediction caused by observing the values of the S subset of attributes for a given instance x. The contribution of the ith feature is defined as its Shapley value with respect to the $v_S(x)$ value function, i.e.:

$$\varphi_i(x) = \sum_{S \subseteq D \setminus \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} \left(v(S \cup \{i\}) - v(S)\right)$$

Figure 6 shows how the features affect the probability of graduation, which is estimated by the neural network. In particular, it shows the distribution of SHAP values for each feature across the students. The figure suggests that high values of *Years elapsed* push the probability of graduation lower, that is probably because students forget a lot during these years. Furthermore, the university entrance score is a valid predictor of academic performance, since the higher the UES is the more likely the student will graduate. The figure also shows that the higher the score of the math matura subexam is, the more likely the students will graduate. Even though BME is a technical university, humanities also play an important role e.g., foreign language and history are important courses since

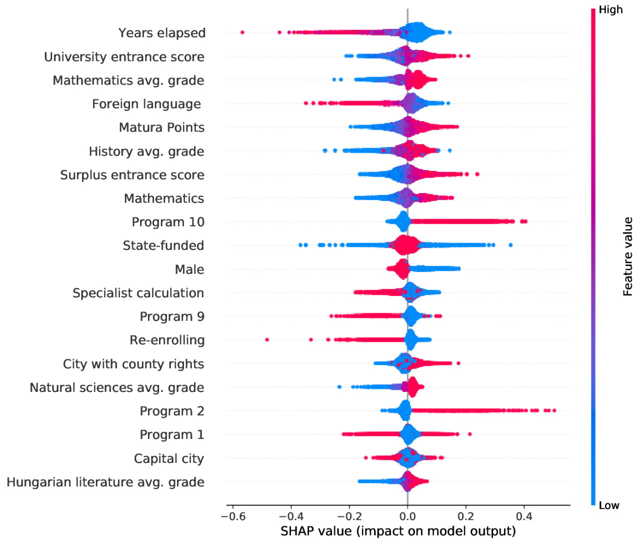

Figure 6: Effects of the features that have a high impact on the probability of graduation. Each student has one dot in each row. The x coordinate of the dot shows the impact of that feature on the prediction, and its color represents the value of that feature for the student.

the scores in these subjects have a high impact on the model's prediction.

The SHAP feature importance is the mean of the absolute Shapley values (average absolute impact on the predicted probability of graduation) of a feature. In Figure 6 the features are ordered according to their importance, i.e. *Years elapsed*, *University entrance score* and *Mathematics avg. grade* are the most important features, the 12 most important features can also be found in Table 4.

6 CONCLUSION

In this work, we used deep neural networks to predict the final academic performance of students at the Budapest University of Technology and Economics with the aim of identifying students at risk of dropping out. Compared to the performance reported in related papers, our deep neural network achieved quite a good accuracy, namely 72.4% (AUC=0.771). Moreover, we also compared the performance of our deep learning model to other cutting edge machine learning algorithms such as XGBoost, and we found that deep learning slightly outperformed the benchmark models, which indicates that deep learning is also suitable for dropout prediction.

To overcome the black-box nature of deep neural networks, we used state-of-the-art model explainability tools to examine how and to what extent the pre-enrollment achievement measures affect university success. We found that the number of years elapsed between the matura exam and enrollment, university entrance score, and mathematics matura sub-exam scores have the highest predictive power on graduation. We can observe that despite the fact that the data come from a technical university, high school performance in humanities also has a high positive impact; this is in line with the fact that generalist students graduate more likely than their peers who focused more on the program-specific matura sub-exams.

As a future step, the prediction accuracy could be potentially increased by experimenting with other state-of-the-art deep learning models, such as Neural Oblivious Decision Ensembles [29], or by taking into account other explanatory variables such as socioeconomic and psychological factors.

ACKNOWLEDGMENTS

We are grateful to Mihály Szabó, Bálint Csabay, and István Bognár from the Central Academic Office of BME for providing the data for this research, and also for their assistance in data understanding. Moreover, we would also like to express our gratitude to Bálint Gyires-Tóth for his suggestions on our research plan, and Tekla Kiss for her help in data preparation. The publication is supported by the EFOP-3.6.2-16-2017-00015 project entitled 'Deepening the activities of HU-MATHS-IN, the Hungarian Service Network for Mathematics in Industry and Innovations'.

REFERENCES

[1] Ghadeer S Abu-Oda and Alaa M El-Halees. 2015. Data mining in higher education: university student dropout case study. *International Journal of Data Mining & Knowledge Management Process* 5, 1 (2015), 15.

[2] Francesco Agrusti, Mauro Mezzini, and Gianmarco Bonavolontà. 2020. Deep learning approach for predicting university dropout: a case study at Roma Tre University. *Journal of e-Learning and Knowledge Society* 16, 1 (2020), 44–54.

[3] R. Alkhasawneh and R. Hobson. 2011. Modeling student retention in science and engineering disciplines using neural networks. In *Proceedings of the 2011 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, Amman, Jordan, 660–663.

[4] Sercan O. Arik and Tomas Pfister. 2019. TabNet: Attentive Interpretable Tabular Learning. arXiv:1908.07442 [cs.LG]

[5] Raheela Asif, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. 2017. Analyzing undergraduate students' performance using educational data mining. *Computers & Education* 113 (2017), 177–194.

[6] Lovenoor Aulck, Dev Nambi, Nishant Velagapudi, Joshua Blumenstock, and Jevin West. 2019. Mining University Registrar Records to Predict First-Year Undergraduate Attrition. In *Proceedings of the 12th International Conference on Educational Data Mining*. EDM, Montréal, Canada, 9–18.

[7] Cédric Beaulac and Jeffrey S Rosenthal. 2019. Predicting university Students' academic success and major using random forests. *Research in Higher Education* 60, 7 (2019), 1048–1064.

[8] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[9] Gerben Dekker, Mykola Pechenizkiy, and Jan Vleeshouwers. 2009. Predicting Students Drop Out: A Case Study. In *Proceedings of the 2nd International Conference on Educational Data Mining*. EDM, Cordoba, Spain, 41–50.

[10] Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan. 2017. A systematic review on educational data mining. *IEEE Access* 5 (2017), 15991–16005.

[11] David S Fike and Renea Fike. 2008. Predictors of first-year student retention in the community college. *Community College Review* 36, 2 (2008), 68–88.

[12] Organisation for Economic Co-operation and Development Staff. 2013. *Education at a glance: OECD indicators 2013*. OECD Publishing, Paris, France.

[13] Antonio Hernández-Blanco, Boris Herrera-Flores, David Tomás, and Borja Navarro-Colorado. 2019. A Systematic Review of Deep Learning Approaches to Educational Data Mining. *Complexity* Special Issue (May 2019), 1–22.

[14] Ulrich Heublein. 2014. Student drop-out from German higher education institutions. *European Journal of Education* 49, 4 (2014), 497–513.

[15] Paul T. Von Hippel and Alvaro Hofflinger. 2020. The data revolution comes to higher education: identifying students at risk of dropout in Chile. *Journal of Higher Education Policy and Management* 0, 0 (2020), 1–22.

[16] Byung-Hak Kim, Ethan Vizitei, and Varun Ganapathi. 2018. GritNet: Student Performance Prediction with Deep Learning. arXiv:1804.07405 [cs.LG]

[17] Botond Kiss, Marcell Nagy, Roland Molontay, and Csabay Bálint. 2019. Predicting Dropout Using High School and First-semester Academic Achievement Measures. In *Proceedings of the 17th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, High Tatras, Slovakia, 383–389.

[18] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. 2017. Self-Normalizing Neural Networks. arXiv:1706.02515 [cs.LG]

[19] Mukesh Kumar, AJ Singh, and Disha Handa. 2017. Literature Survey on Educational Dropout Prediction. *International Journal of Education and Management*

Engineering 2 (2017), 8–19.
[20] Jong Yih Kuo, Chia Wei Pan, and Baiying Lei. 2017. Using Stacked Denoising Autoencoder for the Student Dropout Prediction. In *Proceedings of the 2017 IEEE International Symposium on Multimedia (ISM)*. IEEE, Taichung, Taiwan, 483–488.
[21] A Latif, AI Choudhary, and AA Hammayun. 2015. Economic effects of student dropouts: A comparative study. *Journal of Global Economics* 3 (2015), 4. Issue 2.
[22] Scott Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874 [cs.AI]
[23] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education* 53, 3 (2009), 950–965.
[24] Cindi Mason, Janet Twomey, David Wright, and Lawrence Whitman. 2018. Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. *Research in Higher Education* 59, 3 (2018), 382–400.
[25] Christoph Molnar. 2020. *Interpretable Machine Learning*. Leanpub, Victoria, Canada. https://christophm.github.io/interpretable-ml-book/.
[26] Marcell Nagy and Roland Molontay. 2018. Predicting Dropout in Higher Education Based on Secondary School Performance. In *Proceedings of the 22nd International Conference on Intelligent Engineering Systems*. IEEE, Las Palmas de Gran Canaria, Spain, 389–394.
[27] Marcell Nagy, Roland Molontay, and Mihály Szabó. 2019. A Web Application for Predicting Academic Performance and Identifying the Contributing Factors. In *Proceedings of the 47th Annual Conference of SEFI*. Budapest University of Technology and Economics, Budapest, Hungary, 1794–1806.
[28] Mark Plagge. 2013. Using artificial neural networks to predict first-year traditional students second year retention rates. In *Proceedings of the 51st ACM Southeast Conference*. ACM, New York, NY, United States, 5.
[29] Sergei Popov, Stanislav Morozov, and Artem Babenko. 2019. Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data. arXiv:1909.06312 [cs.LG]
[30] Max Pumperla. 2019. Hyperas. https://github.com/maxpumperla/hyperas.
[31] Neal Raisman. 2013. *The Cost of College Attrition at Four-Year Colleges & Universities. Policy Perspectives.* Technical Report. Educational Policy Institute.
[32] S Ranjeeth, TP Latchoumi, and P Victer Paul. 2020. A Survey on Predictive Models of Learning Analytics. *Procedia Computer Science* 167 (2020), 37–46.
[33] Juan L Rastrollo-Guerrero, Juan A Gómez-Pulido, and Arturo Durán-Domínguez. 2020. Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Applied Sciences* 10, 3 (2020), 1042.
[34] KP Shaleena and Shaiju Paul. 2015. Data mining techniques for predicting student performance. In *Proceedings of the 2015 IEEE International Conference on Engineering and Technology (ICETECH)*. IEEE, Coimbatore, India, 1–3.
[35] Lloyd S Shapley. 1953. A value for n-person games. *Contributions to the Theory of Games* 2, 28 (1953), 307–317.
[36] Ira Shavitt and Eran Segal. 2018. Regularization Learning Networks: Deep Learning for Tabular Datasets. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 1386–1396.
[37] Erik Štrumbelj and Igor Kononenko. 2014. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems* 41, 3 (2014), 647–665.
[38] Johan J Vossensteyn, Andrea Kottmann, Benjamin WA Jongbloed, Franciscus Kaiser, Leon Cremonini, Bjorn Stensaker, Elisabeth Hovdhaugen, and Sabine Wollscheid. 2015. *Dropout and completion in higher education in Europe: Main report*. Technical Report. CHEPS and NIFU.
[39] Wanli Xing and Dongping Du. 2019. Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research* 57, 3 (2019), 547–570.