Implementing a Machine Learning Approach to Predicting Students' Academic Outcomes

Svyatoslav Oreshin
ITMO University
St. Petersburg, Russian Federation
aqice26@gmail.com

Andrey Filchenkov
ITMO University
St. Petersburg, Russian Federation
aaafil@mail.ru

Polina Petrusha
ITMO University
St. Petersburg, Russian Federation
pgpetrusha@itmo.ru

Egor Krasheninnikov
ITMO University
St. Petersburg, Russian Federation
krasheninnikovegor@gmail.com

Alexander Panfilov
ITMO University
St. Petersburg, Russian Federation
sasha_pusha@mail.ru

Igor Glukhov
ITMO University
St. Petersburg, Russian Federation
gluhov2000@gmail.com

Yulia Kaliberda
ITMO University
St. Petersburg, Russian Federation
julia.kaliberda@gmail.com

Daniil Masalskiy
ITMO University
St. Petersburg, Russian Federation
gnom21345@gmail.com

Alexey Serdyukov
ITMO University
St. Petersburg, Russian Federation
leshaserdyukov@gmail.com

Vladimir Kazakovtsev
ITMO University
St. Petersburg, Russian Federation
vokz@bk.ru

Maksim Khlopotov
ITMO University
St. Petersburg, Russian Federation
khlopotov@itmo.ru

Timofey Podolenchuk
ITMO University
St. Petersburg, Russian Federation
p.timon17@gmail.com

Ivan Smetannikov
ITMO University
St. Petersburg, Russian Federation
ismetannikov@itmo.ru

Daria Kozlova
ITMO University
St. Petersburg, Russian Federation
dkozlova@itmo.ru

ABSTRACT

This research is dedicated to the problem of transforming "linear" educational systems of higher education institutions into a new paradigm of person-centered, blended and individual education. This paper investigates role, application, and challenges of applying AI to predict the academic performance traditional of students: dropouts, GPA, publication activity and other indicators to decrease dropouts and make the learning process more personalized and adaptive. In the first part, we overview the process of data mining using internal university's resources (LMS and other systems) and open source data from students' social networks. Such an aggregation allows describing each student by socio-demographic and psychometric features. Further, we demonstrate how we can dynamically monitor students' activities during the learning process to supplement the resulting features. In the second part of our research, we propose various static and dynamic targets for predictive models and demonstrate the results of predictions and comparisons of several predictive models. The research is based on the information on data processing of more than 20000 students in 2013-2019.

CCRIS 2020, Ocotber 27–29, 2020, Xiamen, China

ACM ISBN 978-1-4503-8805-4/20/10…$15.00
https://doi.org/10.1145/3437802.3437816

CCS CONCEPTS

• **Applied computing** → **E-learning**; **Education**; • **Theory of computation** → *Machine learning theory*.

KEYWORDS

Artificial Intelligence, Educational Analytics, Learning analytics, Smart University

ACM Reference Format:
Svyatoslav Oreshin, Andrey Filchenkov, Polina Petrusha, Egor Krasheninnikov, Alexander Panfilov, Igor Glukhov, Yulia Kaliberda, Daniil Masalskiy, Alexey Serdyukov, Vladimir Kazakovtsev, Maksim Khlopotov, Timofey Podolenchuk, Ivan Smetannikov, and Daria Kozlova. 2020. Implementing a Machine Learning Approach to Predicting Students' Academic Outcomes. In *2020 International Conference on Control, Robotics and Intelligent System, Ocotber 27–29, 2020, Xiamen, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3437802.3437816

1 INTRODUCTION

Digital natives [19], the generation that studies now at the Universities require a fundamentally new approach to the process of delivering education. Experts in googling, finding any information in several clicks, they find the traditional, linear approach to educational process ineffective and even boring [2, 18]. They need a flexible, adaptive and interactive learning environment that easily modifies according to the changing situation on the market of human capital, demand to the specific competencies and, that is even more important, to the needs and plans of the digital natives themselves. Such a challenging change in the educational paradigm requires data-driven decisions and powerful and adaptive analytical instruments to handle new problems. Artificial Intelligence is highly likely to be such an instrument because it can be a tool to provide analysis based on large amounts of accumulated data at universities using sophisticated data processing technologies by training predictive machine learning models. Accumulated data helps to predict and analyze learning outcomes for each student and build a personalized educational track based on multiple factors and reach higher educational and personal students' outcomes. Transformation to a new person-centered paradigm of education is proceeding by building services that allow monitoring students' activity and make a decision of personalized impact in order to reduce the likelihood of dropout and improve the learning process.

Nowadays, Learning Management Systems (LMS) have collected a large amount of data of students' performance and can be used to develop Learning Analytics (LA) systems that make it possible to create innovative approaches to increase the effectiveness of overall studying in higher education institutions (HEIs) [15, 21]. These approaches are often called Educational Data Mining (EDM) [21] and Educational Big Data (EBD) [5]. The main difference between these approaches is the level of performance that are analyzed. In this paper, we analyze overall students' performance based on his or her performance in the University with socio-demographic and psychometric data using different parts of approaches mentioned above [5, 15, 21]. Also, the problem of developing and implementing LA [15, 21] solutions in current HEIs educational paradigm are one of the most popular topics today and some of the successes have been done [4, 11]. Almost all the approaches are related with Artificial Intelligence (AI) because it allows to build predictive analytics and other statistical analysis of collected data to build such systems as personalized educational tracks that allow HEI to be more adaptive for a certain student. Also, E-portfolio [8] of a student can be beneficial for further student's career and make student's learning achievements more structured and conceptualised.

In the 2nd chapter we declare and overview the problem of our research. The following chapter is about data-mining process of both internal University data and data from open sources that we use to accumulate enough information about a certain student to fit machine learning models on. After that we discuss the approaches of predicting integral targets (that do not imply any data of student's performance in the University which represent the situation of entering the University) and dynamic targets (using additional data of student's performance in the University and building predictions on targets that change every semester of studying). Finally, we analyze received results and propose future goals for this research.

2 TARGET AND GOALS

This paper investigates the target, goals, methods, features, and challenges of the first, background stage of establishing the integrated educational environment: the development of individual digital portraits of every student in university using analysis of socio-demographic, psychometric, and educational data obtained from university's learning management systems (LMS) and social networks. All these steps are necessary to build a person-centered individual educational track that will be based on student's abilities to study new material, interests, expected outcomes, already gained skills and knowledge.

We discuss all the sources that can be used to collect data about student's performance and such methodology for feature engineering that will be able to use as much information as possible to fit machine learning models. After all the data is accumulated, we demonstrate how this data can be used to predict such targets as student's dropout, student's involvement in scientific activities, getting the first-honor degree, etc. We present and compare the performance of several machine learning models predicting these targets, discuss different ways to measure the results of prediction and to use the models for making a further personalized impact on the obtained group of students. Also, we analyze feature importance to figure out the most important characteristics of a student in predicting different targets.

3 DATA PROCESSING

In this Section, we describe a methodology of developing a unified and comprehensive dataset of each student (digital portrait) in ITMO University using analysis of socio-demographic, educational and psychometric data received from University LMS and social networks.

University of Information Technologies, Mechanics and Optics (ITMO University) is a national research university founded in 1900 and the one of the leading higher education institutions in Russia, providing training and research in advanced science, humanities, engineering, and technology. Offering a wide range of undergraduate, graduate and postgraduate (PhD) programs, the University is home to over 12,000 students, 19% of whom are international students from 87 different countries. ITMO University is continuously holding positions in international academic rankings: in 2019, the University has been ranked 74 worldwide in Times Higher Education (THE) subject rankings in Computer Science, and 76-100 in Automation & Control within Shanghai Ranking by subject.

We used data from the University's internal IT and LMS services to obtain information about students. From these internal sources, we gathered organizational, educational and administrative data as well as educational information such as grades and research activity. As an external source of additional data about students' psychometric characteristics, we used the largest social network in Russia, VK. VK like most social networks core functionality is based around private messaging and sharing photos, status updates and links with friends. According to VK official stats the network counts more than 586000000 registered accounts.

3.1 University Internal Data Processing

In our research, we used three sources of data: Information System of University (ISU), Distance Learning Center (DLC) and social network. DLC collects data about academic performance such as marks, dates of exams, etc. The other information about the student is stored by ISU. Used data is constantly updated which allows us to monitor students' activities dynamically and change predictions according to new obtained information.

The first step of developing students' digital portraits was splitting students according to their study levels (bachelors, masters, etc.). Then we added all the basic features, which contain the known information about students at the time of their enrollment. In order to do this, we needed to obtain basic information, added applicants' information and merged our data with enrollment details for each student. At this point, we have only extracted integral features, that are properties containing information about a student after her/his enrollment without any statistics on attendance and grades. Features created during internal data processing can be divided into three groups:

- Socio-demographic: name, hometown, citizenship, additional fundings for student
- Related to achievements: subject competition prizes, entry or state exams
- Educational: faculty, selected curriculum, group number, student id, free of charge or self-funded education

As a result, we have extracted 55 integral features. The dataframe, which we get in a resulting feature space is an integral digital portrait, which is a description of a student that characterizes him or her at the moment of enrollment in University. Having such a portrait gives us an opportunity to create models that will be able to estimate the probability that the student will successfully finish the University or drop out even before the moment of studies beginning.

In the second step, we added dynamic features to the existing digital portraits. For each semester we appended administrative, educational and research activity data. Extracted features can be divided into three groups:

- Socio-demographic: name, hometown, citizenship, additional fundings for student
- Educational: mean marks, number of academic failures and their standard deviation
- Related to research activities: number of publications indexed in Scopus and other publication databases

These features reflect changes and tendencies in a student's performance over time. It allows us to create and test hypotheses on predicting various target features for each semester. This way we obtain a set of features for each previous semester which can be used as features in our prediction model. For bachelors, the total number of added dynamic features for 8 semesters is 30.

3.2 External Data Processing

As a data source, we choose the most influential social network in Russia, VK (vk.com). It is quite popular among students, thus it is a broad datasource to collect various types of open information about students' interests and behavior.

The first problem we encountered is how we can match students with their social network accounts. Specifically, it is necessary to choose known data from the University correctly for further mapping. Our solution for this problem is to use such data as a full name and birth date to find the account of a certain student in the social network using open data. As a result, we have found pages for 40% of students from ITMO University. This is an acceptable number considering that:

- Not every student has a page in the social network
- Not every student publishes the real name or birth date
- Not every student has an open profile (in VK, a user may choose to make his/her profile private, all the information will be visible only to the friends of the person)

After performing the matching, we were able to obtain all open information from every matched profile to describe its owner to generate features that may represent the social behavior of a particular student.

We had a specific interest in students' personal pages because of a considerable amount of information that was published thereon. Based on the data we gathered from students' pages, we calculated statistics such as the mean number of likes on each page, the total number of posts on a particular page, the mean number of media attached to student's posts, the number of friends, etc. Thus, we collected 14 different attributes. Besides, we divided every calculated statistic into 8 features, one per semester of the study period to avoid data leaks (that is sharing information between train and test datasets, which can affect reliability of results) from the future when making a training set. As a result, we received 112 features in total.

Because of the format of communications in social networks, a big amount of collected data contains texts that include different topics and written in different styles. Thus, the next thing we did was obtaining the text from all posts and reposts of every student's page and applying a topic modeling approach. Our choice for topic modeling algorithms is LDA (Latent Dirichlet Allocation) [7], which allowed us to recover the distribution of topics for each post. Then the bag-of-words technique was applied to get a feature vector for each student describing topics' occurrence and their interests.

One of the things that we are also interested in is the sentiment of opinions in students' posts. For this purpose, we used a pre-trained model for sentiment analysis [1] (Dostoevsky implementation that uses FastTextSocialNetworkModel algorithm). As a result, for each post, we had 5 scores describing its emotional component. Then, for each student, we take a mean vector of all their posts.

As a result, we transformed given data into 55 integral and up to 30 dynamic features from internal sources, as well as 135 features from social networks.

4 PREDICTIVE MODELING

In accordance to two different types of features presented in the student portrait (integral and dynamic), we split possible model targets in two groups as well: static targets, prediction of which would possibly allow us to identify talented or potentially failing students and influence them right after the moment of enrollment, and dynamic targets, prediction of which became more accurate taking into account the dynamic features of the portrait and could help to identify potential problems in the learning process in the nearest future.

4.1 Predictive modeling with static targets

We used the extracted static features to predict the student's future performance which might influence an early stage of the student's education at the University. For example, we can highlight such relevant targets as student's dropout, future scientific activities and the fact of future graduation with the first-honor degree. We can use an output model's probability to obtain probabilities to use it for ranking students according to the chosen target.

Portraits features preprocessing was performed separately for each tested model. For linear classifiers we removed samples with unknown values in features and performed one-hot encoding for categorical features [9]. For other tree-based models unknown values in features were filled with values not presented in feature's domain of definition. Model performance was tested on nested time-series splitted cross-validation [3] which helps to recreate the inference process of these models and calculate quality metrics on previously unknown data.

After cross-validated testing of several predictive models [6, 10, 13] (Table 1), we opted on gradient boosting on decision trees using mean encodings for categorical features [17], which are widely presented in students' portrait features as hometown, gender and etc. The final main metrics results for model performance with confidence interval [20] on chosen targets could be found in Table 2.

As we can see, Catboost outperforms other models due to the big number of categorical features presented in the integral student profile. The ratio of graduate students to dropped is equaled 3 / 5 and observed model results in Table 1 show significance of composed digital profile and proofs the ability of precise different student groups distinguishing with respect to changing target probability threshold. The most important features [22] according to the Catboost model in dropout prediction are: Type of certificate of secondary education (importance is 41%), Hometown (12%), Faculty (8%), Tuition fee funding (7%), Special enrollment condition (6%), Applicant's age (6%), Russian language state exam (3%), Mathematics entry exam (2%), Benefit recipients enrollment (2%), Citizenship (2%).

4.2 Predictive modeling with dynamic targets

First, we extracted students' characteristics for 8 consecutive semesters from the existing datasets. We considered four different targets per student related to students' efficiency respectively:

- Dropout that is the fact of student expulsion from university
- Average mark for all subjects
- Academic failure that is a binary value representing whether student has any academic failures or not
- Average number of academic failures

The data is processed in such a way that neither the target semesters nor subsequent ones were presented in the training set preventing data leakage. We achieved this by creating 32 different training datasets, one per each semester and each target. CatBoost implementation of gradient boosting on decision trees [16] was chosen as a training model, because it provides stable convergence without data normalization and does not require categorical features preprocessing, which is the problem for many other predictive models. Furthermore, at the stage of working with static targets, we revealed that CatBoost implementation outperformed other models, which also served as a sufficient condition for its application. We used the two most representative metrics for binary targets evaluation: recall and ROC-AUC [10]. This decision was motivated by the imbalance of target classes and the necessity of obtaining a degree of prediction confidence, which later would be used for supporting decisions of facilitating selected students. It is more important to find as many risky students as possible and motivate them, this is why we focus on recall. For regression tasks models' quality was assessed with normalized RMSE [14]:

$$nRMSE = \frac{RMSE}{\bar{Y}}$$

where RMSE is the classical root mean squared error and is the mean of all variables in the validation target.

After training and validation, we found out that the majority of models' metrics are commonly increasing from first to last semester because of the additional data and different statistics volume. It can be seen in Table 3 "mean number of academic failures" column. However, this tendency is unstable because of differences in corresponding students' values in semesters.

Table 3 shows the results of fitting for two targets: the average of students' marks in the semester and the average number of academic failures. Average mark prediction shows a poor tendency of improving from semester to semester because all students have their peculiar properties and can demonstrate great results on one set of subjects and bad on another. Furthermore, the subjects are being changed each semester as well as the main direction and difficulty.

First, we extracted students' characteristics for 8 consecutive semesters from the existing datasets. We considered four different targets per student related to students' efficiency respectively:

Table 4 presents the results of the prediction of students' academic failures. Classification for the first two semesters does not show good quality in case of performance data leakage—there are not enough values for building statistics fully describing student tendencies. From the 3rd semester, both the growth of both metrics can be observed as we gain more data about each student with each subsequent semester. This target is the main alarm when it is necessary to define students at risk [15].

Table 5 presents information about the model quality with the corresponding dropout target. We can outline that the recall, in this case, is worse than in case of academic failure prediction. It can be explained with a weak generalization ability of expelled students' signs. They have different reasons and there are not many examples of such students. However, predictions are undoubtedly useful when we want to improve academic performance and avoid unpredictable expulsions.

It is important to define the most important features while analyzing the prediction results. As the most impressive results were shown in Table 4, the layout of features after fitting will be reflected precisely in this model. By "important features" we consider a score

Table 1: Models Performance Comparison On Dropout Prediction With A 95% Confidence Interval

Model	Accuracy	ROC-AUC	Recall	Precission
Logistic Regression	0.84 ± 0.02	0.93 ± 0.01	0.82 ± 0.05	0.71 ± 0.04
Random Forest	0.88 ± 0.01	0.95 ± 0.01	0.84 ± 0.03	0.81 ± 0.02
XGBoost	0.88 ± 0.02	0.96 ± 0.01	0.81 ± 0.04	0.81 ± 0.03
Catboost	0.91 ± 0.02	0.97 ± 0.01	0.83 ± 0.02	0.86 ± 0.03

Table 2: Achieved metric scores on different targets with the Catboost model

Target	Accuracy	ROC-AUC	Recall	Precission
Dropout	0.91	0.97	0.83	0.86
Scientific activity	0.86	0.87	0.86	0.57
First-honor degree	0.94	0.90	0.94	0.41

Table 3: Models' results in average characteristics prediction

Semester	Average GPA (nRMSE)	Number of academic failures (nRMSE)
1	0.121 ± 0.002	1.44 ± 0.04
2	0.095 ± 0.003	1.76 ± 0.07
3	0.113 ± 0.004	1.17 ± 0.06
4	0.091 ± 0.001	1.21 ± 0.14
5	0.104 ± 0.001	1.07 ± 0.05
6	0.086 ± 0.003	1.09 ± 0.05
7	0.095 ± 0.003	0.96 ± 0.03

Table 4: Results in students' academic failures predictions

Semester	Recall	Precission	F1	ROC-AUC
1	0.85 ± 0.03	0.58 ± 0.01	0.69 ± 0.02	0.78 ± 0.02
2	0.71 ± 0.05	0.58 ± 0.01	0.64 ± 0.02	0.783 ± 0.001
3	0.92 ± 0.04	0.67 ± 0.01	0.773 ± 0.015	0.818 ± 0.005
4	0.78 ± 0.07	0.70 ± 0.02	0.74 ± 0.02	0.85 ± 0.01
5	0.92 ± 0.04	0.66 ± 0.01	0.771 ± 0.008	0.81 ± 0.01
6	0.82 ± 0.01	0.76 ± 0.02	0.79 ± 0.01	0.90 ± 0.01
7	0.85 ± 0.05	0.74 ± 0.02	0.792 ± 0.007	0.84 ± 0.02

Table 5: Results in students' dropout predictions

Semester	Average GPA (nRMSE)	Number of academic failures (nRMSE)
1	0.0335	0.833
2	0.429	0.914
3	0.678	0.97
4	0.577	0.952
5	0.489	0.966
6	0.698	0.957
7	0.514	0.99
8	0.613	0.981

that indicates how useful or valuable each feature was in the construction of the boosted decision trees within the model [12]. The most important features are:

(1) Id of student's group (13.9%)
(2) Amount of academic failures in the previous semester (9.9%)
(3) Mean mark in the second semester (9.7%)
(4) Amount of academic failures in the first semester (8.8%)
(5) Faculty (4.1%)
(6) Budget form of education (no tuition fee) (3.9%)
(7) Amount of academic failures in fifth semester (3.5%)
(8) The mean of all academic failures obtained till current (2.9%)
(9) Common enrollment condition (2.8%)
(10) Amount of academic failures in fourth semester (2.6%)

As we can see, the most important features are related with previous student's experience having academic failures and average marks obtained during the studying process up to the current moment of prediction. Also features such as group number, which contains information about students' educational track and a group level, faculty, format of education and enrollment conditions are significantly affecting the model's quality.

Feature importances of models that are fitted on dynamic targets do not differ much, so feature importance obtained from the model fitted on the 7th semester of academic failure prediction is a representative example.

5 RESULTS AND DISCUSSIONS

In our research, we achieved high results in predicting both integral and dynamic targets. This was able due to rigorous and comprehensive data processing and feature engineering. We used all the possible internal and external data sources. We will continue our research by collecting more data to increase performance of predictive models and by using these models to make advanced analytics and special personalized impact on students who have difficulties with studying. Such a monitoring system would allow running AB-tests to figure out the most effective way to decrease a dropout rate and increase the learning outcomes of students and make them more satisfied with studying in the University.

An interesting and simultaneously expected result is that the quality of prediction growth with the increasing number of semesters. This is not only due to we use more semesters to predict the results of the later semester, but also we think that it is due to the students' learning strategies get more stable, so students who cares about the result fail less often than in starting semester.

Also, the obtained results of predicting different targets clue that it is possible to make a recommendation system using this data with a unified description of academic disciplines, MOOCs, internships, and other learning activities. Such a recommendation system would allow a University building personalized learning tracks and graduate more specifically qualified professionals. This would increase students' involvement in their learning process and motivate them to get the exact skills and qualifications they want.

6 CONCLUSION

Analyzing the results of our research, we can conclude that it is possible to make high-quality predictions for the proposed targets using machine learning. These achievements indicate that collecting a unified representation of a student using data from different sources and aggregating it in a correct way allows predicting targets such as the probability of dropouts, scientific outcomes and getting the first honor using static features that are obtained only since the moment when a student enters the University. Also, we reviewed and discussed an approach for collecting data and predicting dynamic targets by adding newly gained information about student's performance at the University. We demonstrated the results of dynamic targets prediction in the following semester as the probability of student's dropout and getting an academic failure and continuous targets as an average GPA and number of academic failures.

As the main goal of the research, we indicated the problem of the digital transformation of higher education institutes and proposed an approach to collecting and processing accumulated data. The introduction of such a monitoring system can optimize many internal University processes and would allow helping students in building and correcting their personal tracks, making the educational process more adaptive, integrated and matching students' needs and demands on the human resource market. Also, digital transformation can make a university and students more connected with each other by using a person-centered paradigm of higher education that can be archived by applying artificial intelligence technologies into the modern educational system.

REFERENCES

[1] A.Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media. Association for Computational Linguistics.* 30–38.

[2] R. S. J. Baker. 2010. Data mining for education. *International encyclopedia of education, Oxford* 7, 3 (2010), 112–118.

[3] C. Bergmeir and J. M. Benítez. 2012. On the use of cross-validation for time series predictor evaluation. *Information Sciences* 191 (2012), 192–213.

[4] R. Bose and J. Frew. 2005. Lineage retrieval for scientific data processing: A survey. In *ACM Computing Surveys (CSUR).* 1–28.

[5] S. Brull and S. Finlayson. 2016. Importance of gamification in increasing learning. *Journal of Continuing Education in Nursing* 47, 8 (2016), 112–118.

[6] T. Chen and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Processings of The 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 785–794.

[7] B. M. David, Y. Andrew Ng, and J. I. Michael. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 4-5 (2003), 993–1022.

[8] C. Ellis. 2017. *The Importance of E-Portfolios for Effective Student-Facing Learning Analytics.* Springer Nature Singapore Pte Ltd., Chapter 3, 35–49.

[9] D. Harris and S. Harris. 2012. *Digital design and computer architecture (2nd ed.).* Morgan Kaufmann Publishers.

[10] J. D. Kelleher, B. Mac Namee, and A. D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies.* MIT Press Ltd.

[11] S. R. Kummar and S. Hamid. 2017. Analysis of Learning Analytics in Higher Educational Institutions: A Review. In *Lecture Notes in Computer Science.* 185–196.

[12] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts. 2013. Understanding variable importances in Forests of randomized trees. *Advances in Neural Information Processing Systems* 26 (2013).

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[14] A. Poli and M. C. Cirillo. 1993. On the use of the normalized mean square error in evaluating dispersion model performance. *Atmospheric Environment. Part A. General Topics* 27, 15 (1993), 2427–2434.

[15] M. Prensky. 2001. Digital Natives, Digital Immigrants Part 1. *On the Horizon* 9, 5 (2001). https://doi.org/10.1108/10748120110424816

[16] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. 2018. CatBoost: gradient boosting with categorical features support. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018).*

[17] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. 2018. CatBoost: unbiased boosting with categorical features. In *NeurIPS.*

[18] S. Rajeswari and R. Lawrance. 2016. Classification model to predict the learners' academic performance using big data. In *Proceedings of International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16).*

[19] C. Romero and S. Ventura. 2015. 2015. J. A. Larusson, B. White (eds): Learning Analytics: From Research to Practic. *Technology, Knowledge and Learning* 20, 5 (2015), 357–360.

[20] J. L. Snell and S. R. Fisher. 1959. Statistical Methods and Scientific Inference. *Econometrica* 27, 1 (1959), 138.

[21] R. Srivastava, M. Gendy, M. Narayanan, Y. Arun, and J. Singh. 2012. *University of the future—A Thousand Year Old Industry on the Cusp of Profound Change.* Ernst & Young. https://web.archive.org/web/20121119092916/http://www.ey.com/Publication/vwLUAssets/University_of_the_future/%24FILE/University_of_the_future_2012.pdf

[22] Z. Xu, G. Huang, K. Q. Weinberger, and A. X. Zheng. 2014. Gradient Boosted Feature Selection. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14).*