BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **22**, No 1
Sofia • 2022 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2022-0008

Early Student-at-Risk Detection by Current Learning Performance and Learning Behavior Indicators

Tatiana A. Kustitskaya, Alexey A. Kytmanov, Mikhail V. Noskov

School of Space and Information Technology, Siberian Federal University, Krasnoyarsk, Russia
E-mails: tkustitskaya@sfu-kras.ru aakytm@gmail.com mvnoskov@yandex.ru

Abstract**:** *The article is focused on the problem of early prediction of students' learning failures with the purpose of their possible prevention by timely introducing supportive measures. We propose an approach to designing a predictive model for an academic course or module taught in a blended learning format. We introduce certain requirements to predictive models concerning their applicability to the educational process such as interpretability, actionability, and adaptability to a course design. We test three types of classifiers meeting these requirements and choose the one that provides best performance starting from the early stages of the semester, and therefore provides various opportunities to timely support at-risk students. Our empirical studies confirm that the proposed approach is promising for the development of an early warning system in a higher education institution. Such systems can positively influence student retention rates and enhance learning and teaching experience for a long term.*

Keywords**:** *Learning analytics, Learning success, Learning failure, Student-at-risk, Early warning system, Bayesian network, k-Nearest Neighbors, Linear discriminant analysis.*

1. Introduction

The rapid development of technology encourages Higher Educational Institutions (HEIs) to invest significant resources in the digital transformation of the educational process (including shifting towards distance learning techniques). Automation of routine procedures can free up resources and personnel to focus on more innovative and growth-impacting work, such as science-intensive research or developing high-level courses. However, the increasing spread of online learning and blended learning technology poses new challenges to HEIs such as loss of awareness and control over certain components of the learning process.

Fortunately, with existing tools, it is possible to collect a variety of educational data which can provide one with insights into learners' behavior and ways of achieving learning outcomes by means of its thorough analysis [1-3].

All of the above became prerequisites for the emergence of Learning Analytics (LA), a relatively new branch of Data Science that comprises aims and methods drawn from Educational and Psychological studies. According to the most well-known definition, LA is the measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs [4].

LA is expected to provide benefits for all the stakeholders (students, teachers, designers, administrators) of the higher education marketplace [5]. For instance, students may benefit from LA through personalized and adaptive support of their learning journey [6].

A considerable number of students who continuously face demanding educational requirements and challenges of university life become unable to cope successfully with their compulsory educational duties. This leads to numerous dropouts, especially among freshmen [7, 8]. A timely signal of an increased risk of failure in training (e.g., deviation from the expected schedule for mastering the course, changes in educational behavior, and a significant decrease in academic performance) provides an opportunity to implement supportive measures that can return a student to the initially prescribed track and ensure successful completion of the course.

Among such measures there can be introducing personalized tutoring or extracurricular reinforcement courses. As mentioned in [9], the development of a supportive campus environment, incorporating various pedagogical approaches, validation and teaching activities might considerably improve student success rates.

An Early Warning System (EWS) is one of the effective instruments of LA for developing a supportive environment. It uses student data to generate indicators of on-track status for graduation or course completeness, identify students who encounter problems with mastering a particular course and direct them to appropriate interventions [10].

A number of educational institutions have already implemented EWS in their educational process with most of the examples coming from the USA, where they have been under extensive use for about a decade. The most well-known example of EWS is the training success support system “Course Signals” of Purdue University (Indiana, USA). Other examples of such systems are “Grade Performance Status” of Northern Arizona University, (Northern Arizona, USA), Leaning Intelligent System LIS in The Open University of Catalonia (Spain) and others.

As for the Russian universities, they are still on the way to establishing such systems with some elements of them present today at Lomonosov Moscow State University, Higher School of Economics, Saint-Petersburg State University, Moscow Engineering Physics Institute, Bauman Moscow State Technical University, ITMO University, Tyumen State University, and the “Digital University” project implemented by three universities located in Tomsk: Tomsk Polytechnic University, Tomsk State University, and Tomsk State University of Control Systems and Radioelectronics

There are several national initiatives aimed at digitalisation of educations comprising a data-driven approach to educational management and enhancement of the learning environment such as University-2035 and Consortium “Evidence-based digitalization for student success”.

One of the drivers of development in implementing EWS (as well as other LA instruments) in education is the COVID19 pandemic, which has seriously increased the importance and the spread of online learning [11, 12]. At the same time, it has upset the balance between online and face-to-face education, which complicates the interaction between course instructors and learners [13]. This has subsequently increased the need for reliable tools for the timely identification of underperforming students and their further support.

In this paper, we focus on developing and testing a predictive model for learning success or failure. The aim of our pilot study is to build, for a specific course, an accurate predictive model for detecting at-risk students at early stages. We introduce certain requirements to predictive models concerning their applicability to the real educational process. We then test three types of classifiers meeting these requirements with respect to various quality metrics and choose the best one.

2. A review of studies on learning success and failure prediction

There is no universal definition of "learning (academic or study) success" [14]. One reason for this may lie in different perspectives of "success" in the views of students, teaching staff, or society [15]. Such things as a good final grade in a certain course, an acceptable GPA, achievement of a degree, satisfaction with education, employability, development of student's professional competencies might all be considered as criteria for learning success. The relevance of student success prediction is ensured by a large number of publications on the topic. In [16], authors state that, according to their search, over the seven years from 2011 to December 2017, 164 papers on the topic have been published in 46 journals and 33 conference proceedings indexed in Scopus or Web of Science databases.

As learning environments of HEIs vary, researchers focus on data easily available from particular information systems and LMSs, and variables significant as learning success predictors for the specific educational environment. A considerable number of case studies have been devoted to exploring the best-suited predictive algorithms and significant success/failure predictors, as well as to testing the implemented models.

In the works briefly reviewed below, authors suggest a range of models used to predict student performance including quite elaborated predictive models.

We start with works considering classical probabilistic classifiers.

Namely, [17] focuses on developing and applying a Naive Bayesian Classifier to data from a Learning Management System (LMS) to predict the dropout rate, using such criteria of student performance as number of inputs, time spent in the e-Learning environment, weighted number of inputs, and weighted time spent as predictors.

In [18] authors develop a Bayesian Network to predict students' final grades in the course in Mathematics, using such predictors as gender, attitude to teamwork, interest in math, motivation for studying, self-confidence, shyness, the level of English.

In [19] a EWS to identify at-risk students is developed using Logistic Regression built on such key variables as total number of messages posted in forum discussions, total number of email messages sent, and total number of assignments completed.

In [20] and in [21] students are initially split into three groups (unsuccessful, successful at the minimum level, and successful students) as they undergo some initial assessment. A student can be moved between the groups based on their performance throughout the semester, and authors predict the probabilities of student transition from a certain group to another by means of Markov processes and Markov chains.

Another promising class of machine learning approaches are logical models (association rules, decision trees and their ensembles), which mirror a natural way of decision making and do not require much preprocessing of data. In [22, 23] authors apply various types of such algorithms to identify successful and unsuccessful learners using data from e-Learning platforms.

Black-box models are often criticized for their non-transparency, but in case of large amounts of poorly structured data and time-consuming process of feature selection (which is especially typical for MOOCs) researchers opt for deep learning models and achieve an acceptable quality of prediction.

For instance, in [24] authors propose a deep neural network, a combination of convolutional neural networks and recurrent neural networks, which automatically extracts features from raw data, for solving dropout prediction problems in MOOCs. They argue that the proposed model can achieve comparable performance to approaches relying on feature selection performed by experts.

In [25] authors mention the considerable potential of learners' time series data for early prediction of learning outcomes. In their study they use students' daily click frequencies without any other auxiliary information to predict their final course performance, using a deep learning approach to avoid extensive work on feature engineering which is usually required for the accuracy of classical machine learning algorithms.

In most studies, authors either choose between numbers of machine learning models of different types or combine algorithms to heighten the quality of prediction.

In [26] authors combine Logistic Regression, Linear Discriminant Analysis, and Support Vector Machines for the success/failure prediction. They use a number of features available from a learning platform as predictors and conclude that the pace of activities (i.e., the frequency of events) performed by students in the platform used as the only predictor produces the most accurate results.

In [27] there is presented a study the on the effectiveness of various algorithms (Naive Bayes, Classification Tree, Random Forest, Support Vector Machines, Neural Network, CN2 Rules, and k-Nearest Neighbours) for early prediction of student success using data from an online learning environment. The best classification performance has been shown by the k-Nearest Neighbors method and CN2 Rules.

In [28] authors predict student performance in a course using the Orange Data Mining System (**https://orangedatamining.com/**) for building several models of classification (Tree, Random Forest, Logistic Regression, Naive Bayes, Support Vector Machines and Neural Network). Using these models students' performance is classified into five groups (Bad, Middle, Good, Very Good and Excellent) and detection of "students at critical zone" can be considered as one of the resulting achievements of the developed predictive system. The Random Forest algorithm provided the best prognostic accuracy in their case. At the same time, authors mention that combining the results from different machine learning algorithms might result in even higher accuracy.

There are related studies in Siberian Federal University where the educational process utilizes a blended learning approach based on certain principles that constitute altogether the so-called Polyparadigm Approach described in [29]. One of the major principles states that continuous assessment of students' results through electronic learning systems is essential for achieving their learning success. The in-class education is accompanied by the distant work of students in the Moodle-based LMS “E-Courses” (**https://e.sfu-kras.ru/?lang=en**) capable of collecting and storing significant amounts of data on student learning behaviour and learning progress.

For the purpose of administrative management of the educational process at the School of Space and Information Technology, there has been developed an automated management information system “AIS SSIT” (**http://dec.sfu-kras.ru/**) consisting of several independent modules. The “Electronic dean’s office” module, for instance, allows teaching and administrative staff to track student attendance, their current scores in “e-Courses” and final exam grades

These data were used in some earlier studies on learning success prediction. For instance, [30] introduces a comprehensive student rate, which incorporates information about student total scores, attendance, and a number of “effective” entries into an e-course. Authors propose a predictive model for a comprehensive student success rate based on the birth-death process. They use intensities of obtaining and assimilating information as the parameters of the process. The approach looks rather beneficial for the administrative staff because the formula provides a simple indicator of the failure risk and helps to evaluate the overall situation.

At the same time, the existing tools do not cover all the aspects of the educational process, as they do not take into account peculiarities of a particular subject area and a course design, as well as in-class student performance and their engagement in the learning process.

To summarize, a wide range of predictive models is already in use for learning analytics, but it is hardly possible to define a universal design of a predictive model fitting all types of learning environments and different research aims.

In our study, we aim at designing a predictive model for blended learning with the possibility to use it not only for forecasting but also for a better understanding of the learning process. Thus, it is important to define the basic principles for the model designing for this particular aim.

3. An approach to timely student-at-risk detection

We start with formulating the desired properties of a predictive model.

In a blended learning format, grading of a student together with the other assessments of their learning behavior can significantly depend on the course instructor and course design. Thus, the predictive model should provide the ability to adjust to a particular course considering its structure and the significance of various learning indicators for student success. We call this property *adaptability to a course design*.

Sharing the idea of the usefulness of explanatory models for LA [31, 32] we are looking for a model that, along with accuracy, has good *interpretability*. A highly interpretable model provides a better understanding of causal relationships between characteristics of learners and the learning process, or a course structure. Such insights might contribute to course design improvements and cognitive models of learning.

Another desired property for a predictive model is its *actionability* [33], which means that a model should be able to react immediately to the new information about student behavior. This means that predictions should be done by either a dynamic model or a set of static models – each designed for a certain time period. The choice of an option should be, among other things, based on how the collection of student data in a particular educational institution is organized. For example, if in a learning course grades and attendance are assigned manually once a week, then it is sufficient to have one static model per week.

Obviously, the overall good quality of the forecast is important. However, we insist that special attention should be paid to *high predictive performance at early stages* as in the first half of an academic semester the chance to correct the student's learning behavior is much higher.

To make ourselves precise, we define learning success as getting a passing final grade, and learning failure as getting a failing grade. We regard the student as a student-at-risk if a predictive algorithm classifies his/her performance as a learning failure. Thus, the response variable for our model is binary with "1" corresponding to failure and "0" corresponding to no failure (i.e., success).

Classical predictive models for student-at-risk detection. The points described above made us opt for white-box models. For the current study, we chose a K-Nearest Neighbors (KNN) classifier, Linear Discriminant Analysis (LDA) and a classifier Constructed on a Bayesian Network (BNC).

To achieve the goal of early prediction of learning failure, it is necessary to make forecasts from the very beginning of the semester. At the time of forecasting, the values of a number of predictors can still be unknown as the corresponding training activities have not yet been carried out. Accordingly, the forecasting system must take into account which predictor values are known to date, and make predictions only on their basis. This can be done in various ways, for example by building a set of forecast models of the same type.

In our work, we build a set of 17 models based on the KNN algorithm and 17 models based on the LDA algorithm (one model of each type for a week). To account for new information about learners each week, we widen a set of predictors by the scores for later learning activities, attendance, etc.

The Bayesian approach allows us to avoid building a set of models, thanks to the possibility of gradually updating the model by adding new information into it.

A Bayesian network is a directed acyclic graph, whose nodes are random variables, and edges represent conditional dependencies between variables.

In our case, we regard learning failure predictors and the response variable as nodes. Prior conditional probabilities can be obtained by expert judgment or using statistical inference from a dataset which contains the data on performance of the students, who have already got the final exam grades.

To predict the probability of learning failure at a certain point, we need to include into the network new information on the predictor values known up to the moment (i.e., to get evidence). Having done that, we compute the posterior distribution for the response using Bayes' rule, and classify the student as a student-at-risk if his/her posterior probability of failure is larger than a previously defined threshold.

It is worth mentioning that probabilistic implications for any variable (not only the response one) can be readily computed given evidence. Such inferences can be useful for constructing models of a learner, as well as for improving the tutoring system. However, there are multiple contentious issues underlying any causal interpretation of Bayesian networks [34]. Thus, any causal discovery in a network should be treated with caution.

4. Student-at-risk detection in the course of Probability and Statistics

To implement the designed approach of student-at-risk detection in the educational process in the SSIT and to assess its functionality, we used the chosen classifiers to accompany a course of Probability and Statistics.

The course is taught using the technology of blended learning, which means that students are obliged to attend lectures and practicums, and do out-of-class independent work in the corresponding electronic course, completing individual e-Tests. Within the course students get practicum, quiz, e-Test and attendance scores. All assessment tools are integrated into the electronic course.

4.1. Preliminary student performance analysis

We collected the data on student performance and final grades for the course of Probability and Statistics for three consecutive academic years (2016-2019). The dataset was formed by the educational data of the total of 129 students.

At the first stage we conducted a preliminary descriptive and visual analysis of data on students' final grades and their performance throughout the semester.

There are five levels of grades: "2", "3", "4", or "5" for bad, satisfactory, good or excellent performance, respectively, or "n/a" ("not/awarded") grade for not participating in a certain activity or failing to submit an assignment. The distribution of the final grades is the following: "n/a" – 22.3%, "2" – 23.9%, "3" – 16.8%, "4" – 20.8% and "5" – 16.2 %, so the dataset is balanced by the classes "learning success" and "learning failure".

During a semester, students take three quizzes (at weeks 6, 13, and 17). The distribution of their grades is presented in Fig. 1.

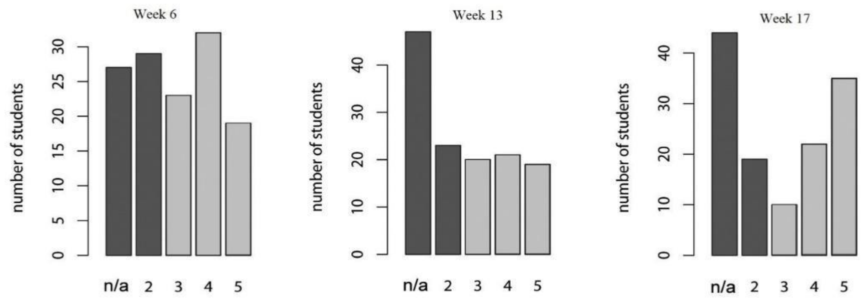

Fig. 1. Distribution of the grades for the quizzes

Grades for the quizzes are quite strongly correlated with the final grades for our dataset (the correlations are 76%, 79%, and 79%, respectively). This allows us to consider the grades for the quizzes as good predictors of learning success.

In addition, students take 16 electronic tests as a part of their individual work during a semester. Correlations between the e-tests scores and the final grades are pretty weak and vary from 24% to 56%.

Together with the e-tests scores, the dataset also contains the information on the number of attempts made by a student to pass a particular test. The maximum allowed number of attempts varies from 4 to 7 depending on the test. Usually, a student makes a small number of attempts (1 or 2). A larger number of attempts indicates, on the one hand, problems with understanding the material, and, on the other hand, a high level of persistence and motivation. Thus, attempts numbers can have a valuable input to learning failure prediction.

Lectures and practicums are held once a week, so that we have a total number of 17 lectures and practicums. Final attendance percentage (available at the end of a semester) correlates quite strongly with final grades (correlations between attendance and final grades are 69% and 74% for lectures and practicums, respectively). However, the intermediate correlations are considerably lower (43% and 48% at week 4; 59% and 64% at week 9), which might indicate that attendance rates are not good predictors of learning failure at early stages.

4.2. Building predictive models for student failure

Classification models include the following indicators of learning performance and behavior, whose values become known at different points in time during a semester:

A binary response variable:

FinalExamFailure – a final exam grade (1 – a student has failed/not awarded the grade at the exam, 0 – a student has passed the exam);

Predictors*:*

Lec – a number of lectures, attended by a student to a certain week;

Prac – a number of practicums, attended by a student to a certain week;

Plus – a number of points awarded for good in-class performance by a certain week;

Quiz$_i$ – a score for the *i*-th quiz, $i=1,\ldots,n$;

e-Test$_i$ – a score for the *i*-th e-Test, $i=1,\ldots,m$;

Persist$_i$ – an indicator of student persistence in their work on the i-th e-Test, i=1,…,m, which depends on the numbers of attempts to pass the e-Test (0 – a student did not attempt to take the test, 1 – a student made 1 attempt, 2 – a student made 2 attempts, 4 – a student used the maximum possible number of attempts, 3 – other cases).

As predictions are made weekly, we built a set of 17 KNN models and a set of 17 LDA models.

The structure of the Bayesian network is the same for all the weeks of a semester (Fig. 2).

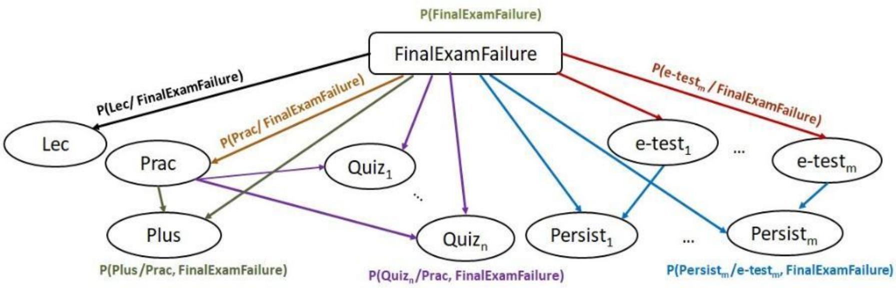

Fig. 2. The Bayesian network structure

The operation of the classifier based of the Bayesian network is organized as follows:

● each week, after obtaining up-to-date data on student performance, the Bayesian network recomputes posterior distributions;

● the posterior probability of the event FinalExamFailure=1 is compared with a threshold p, and, if it exceeds p, the student is classified as an at-risk student for this particular week.

Although Bayesian networks can incorporate evidence at any time, in order to compare the performance of all the considered classifiers, we carry out the described operation once a week.

4.3. Estimation of the model performance

To estimate predictive efficiency of the classifiers, we build predictive models on the same dataset and the same set of predictors using KNN, LDA and BNC. We use K=3 for the KNN algorithm as this value of K provides the best performance of the classifier in our case.

To assess whether the classifiers appear to identify successful and unsuccessful students properly, we use a number of standard classification performance metrics (*accuracy, sensitivity, precision, specificity* and the *weighted F-score*):

$$\text{weighted } F-\text{score}=\frac{(1+\beta^2)\times\text{sensitivity}\times\text{precision}}{\beta^2\times\text{presicion+sensitivity}}, \quad (1)$$

where $0<\beta<1$ prioritizes precision rather than sensitivity, while $\beta>1$ prioritizes sensitivity (and, in our case, gives more attention to detecting as many students at-risk as possible).

In the current problem setting, we regard sensitivity as a much more valuable measure of quality for a classifier than accuracy, precision and specificity since the most important task for the warning system is to detect all the students who are likely to fail the exam.

Nevertheless, the students who perform at a good enough level should not be frequently disturbed with warning messages, and consequently the percentage of true positives should also be taken into consideration.

We therefore set β=2 and regard the weighted F-score as a criterion for choosing the best classifier.

It is reasonable to expect that the efficiency of a built predictive model will increase over time as we get more extensive evidence about student performance. At the same time, the earlier we can detect at-risk students, the more effectively we can bring supportive measures into their study process. Thus, among the chosen models of classification we prefer the one, whose sensitivity and weighted F-score reach acceptable values at earlier stages.

To compare performance of the classifiers, we form fifteen testing sets by randomly mixing the data from the original dataset. On each set we train BNC, KNN and LDA models and estimate the quality of classification using a cross-validation procedure.

The detailed comparison of the models performance is made in [35]. In this article we report on the comparison of the most important metrics in our case – sensitivity and the weigted F-score.

In Table 1, sensitivity of BNC exceeds sensitivity of the other algorithms in the vast majority of cases starting from the very beginning of the semester, and ranges between 0.73 and 1.

Table 1. Sensitivity for BNC, KNN and LDA on 4 out of the 15 testing sets

Week	Set1			Set 2			Set 3			Set 4		
	BNC	KNN	LDA	BNC	KNN	LDA	BNC	KNN	LDA	BNC	KNN	LDA
1	**1.00**	**1.00**	0.38	0.93	**0.94**	0.16	**1.00**	**1.00**	0.22	**1.00**	0.98	0.27
2	**1.00**	0.91	0.47	**1.00**	0.97	0.31	**1.00**	0.98	0.42	**0.97**	0.95	0.47
3	**0.88**	0.85	0.70	0.80	**0.86**	0.59	**0.94**	0.85	0.65	**0.83**	0.76	0.68
4	**0.85**	0.69	0.58	**0.73**	0.67	0.53	**0.83**	0.68	0.53	**0.85**	0.58	0.46
5	**0.77**	0.76	0.67	**0.84**	0.68	0.46	**0.88**	0.60	0.46	**0.90**	0.48	0.50
6	**0.83**	0.74	0.64	**0.86**	0.66	0.51	**0.88**	0.68	0.52	**0.83**	0.54	0.50
7	**0.75**	0.66	0.66	**0.79**	0.54	0.48	**0.83**	0.57	0.63	**0.80**	0.49	0.56
8	**0.78**	0.69	0.72	**0.89**	0.57	0.66	**0.89**	0.65	0.78	**0.74**	0.54	0.67
9	**0.78**	0.69	0.71	**0.87**	0.62	0.67	**0.90**	0.66	0.72	**0.76**	0.64	0.69
10	**0.88**	0.69	0.82	**0.89**	0.61	0.73	**0.98**	0.69	0.73	**0.86**	0.61	0.70
11	**0.85**	0.66	0.78	**0.86**	0.59	0.73	**0.90**	0.70	0.79	**0.90**	0.64	0.72
12	**0.85**	0.66	0.82	**0.86**	0.68	0.76	**0.90**	0.72	0.80	**0.85**	0.66	0.63
13	**0.85**	0.70	0.82	**0.85**	0.64	0.77	**0.92**	0.70	0.77	**0.88**	0.69	0.68
14	**0.85**	0.70	0.87	**0.89**	0.64	0.83	**0.84**	0.74	0.75	**0.86**	0.75	0.79
15	**0.91**	0.72	0.87	**0.78**	0.64	0.76	**0.80**	0.74	0.78	**0.83**	0.72	0.82
16	**0.87**	0.72	0.87	**0.86**	0.64	0.75	**0.85**	0.74	0.75	**0.84**	0.68	0.79
17	**0.85**	0.76	0.81	**0.78**	0.68	0.70	**0.83**	0.74	0.71	**0.89**	0.74	0.78

A similar picture can be oserved in Table 2, where BNC provides the best classification quality in terms of weighted F-score starting from the fourth week of the academic semester (the weighted F-score takes values in the interval [0.69, 0.89]).

Table 2. The weighted F-score for BNC, KNN and LDA on 4 out of the 15 testing sets

Week	Set1			Set 2			Set 3			Set 4		
	BNC	KNN	LDA	BNC	KNN	LDA	BNC	KNN	LDA	BNC	KNN	LDA
1	0.78	**0.97**	0.39	0.75	**0.93**	0.48	0.77	**0.98**	0.32	0.78	**0.99**	0.37
2	0.78	**0.90**	0.55	0.80	**0.95**	0.43	0.77	**0.94**	0.49	0.76	**0.94**	0.49
3	0.78	**0.83**	0.69	0.74	**0.86**	0.58	**0.83**	0.82	0.65	0.74	**0.78**	0.67
4	**0.82**	0.69	0.56	0.70	**0.75**	0.60	**0.76**	0.69	0.61	**0.81**	0.68	0.48
5	**0.75**	**0.75**	0.64	**0.83**	0.69	0.51	**0.83**	0.61	0.54	**0.85**	0.66	0.57
6	**0.81**	0.74	0.62	**0.85**	0.67	0.57	**0.83**	0.69	0.60	**0.80**	0.65	0.58
7	**0.73**	0.67	0.66	**0.78**	0.63	0.63	**0.79**	0.59	0.65	**0.78**	0.59	0.56
8	**0.76**	0.70	0.72	**0.87**	0.67	0.65	**0.85**	0.67	0.78	**0.69**	0.65	0.69
9	**0.77**	0.71	0.71	**0.87**	0.73	0.69	**0.87**	0.69	0.74	**0.73**	0.67	0.71
10	**0.86**	0.70	0.79	**0.88**	0.70	0.72	**0.93**	0.72	0.72	**0.81**	0.63	0.72
11	**0.84**	0.69	0.76	**0.86**	0.69	0.71	**0.86**	0.73	0.77	**0.85**	0.67	0.73
12	**0.84**	0.69	0.78	**0.86**	0.77	0.74	**0.87**	0.74	0.79	**0.81**	0.68	0.74
13	**0.84**	0.71	0.78	**0.86**	0.73	0.75	**0.87**	0.73	0.76	**0.84**	0.72	0.69
14	**0.85**	0.72	0.83	**0.89**	0.73	0.81	**0.82**	0.76	0.75	**0.83**	0.76	0.80
15	**0.89**	0.75	0.83	**0.78**	0.73	0.75	0.75	0.76	**0.77**	**0.80**	0.74	**0.80**
16	**0.86**	0.74	0.82	**0.86**	0.73	0.74	**0.82**	0.76	0.75	**0.79**	0.70	0.77
17	**0.82**	0.76	0.78	**0.79**	0.78	0.69	**0.78**	0.75	0.71	**0.84**	0.76	0.78

The average values of sensitivity and the weighted F-score over the fifteen testing sets have been calculated for BNC, KNN and LDA algorithms. We compare the calculated values by plotting the excess (difference) of the values of the corresponding metrics of the BNC model compared to the KNN model (Fig. 3) and that of the BNC model compared to the LDA model (Fig. 4).

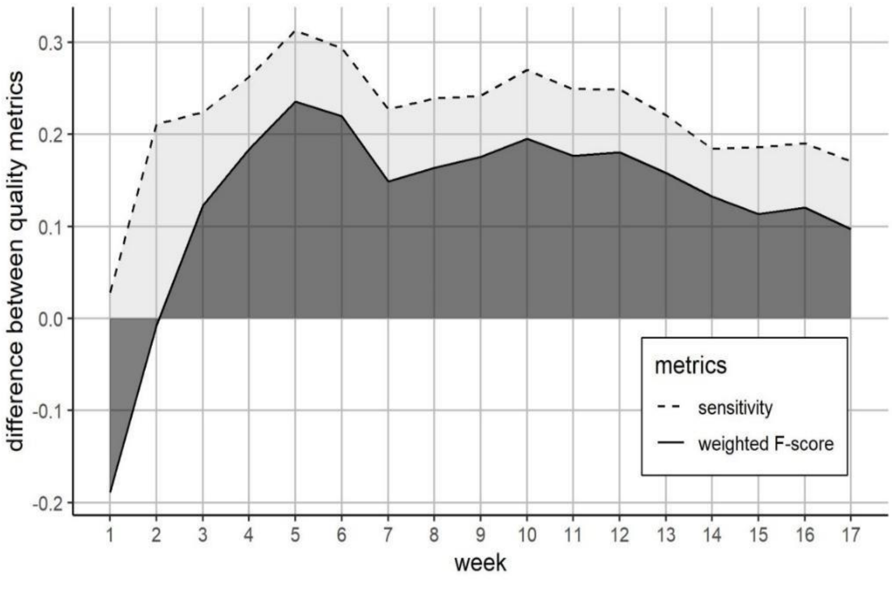

Fig. 3. Average difference of classification performance metrics for BNC and KNN

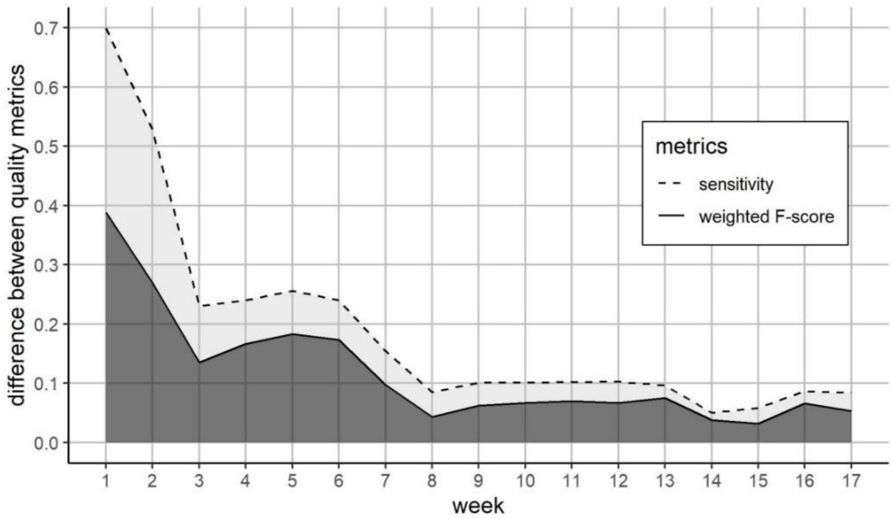

Fig. 4. Average difference of classification performance metrics for BNC and LDA

Since according to the target quality metrics, BNC shows the best efficiency at the early stages, it has been chosen as a predictive model for a EWS for the course.

4.4. Testing the model on new data and developing a feedback mechanism

The empirical study on testing the developed predictive model and implementing a partially automated feedback mechanism was conducted in 2020 on the group of 75 students. Testing the predictive performance of the BNC on these new data, we found that the quality of forecast has improved for all the metrics (Fig. 5).

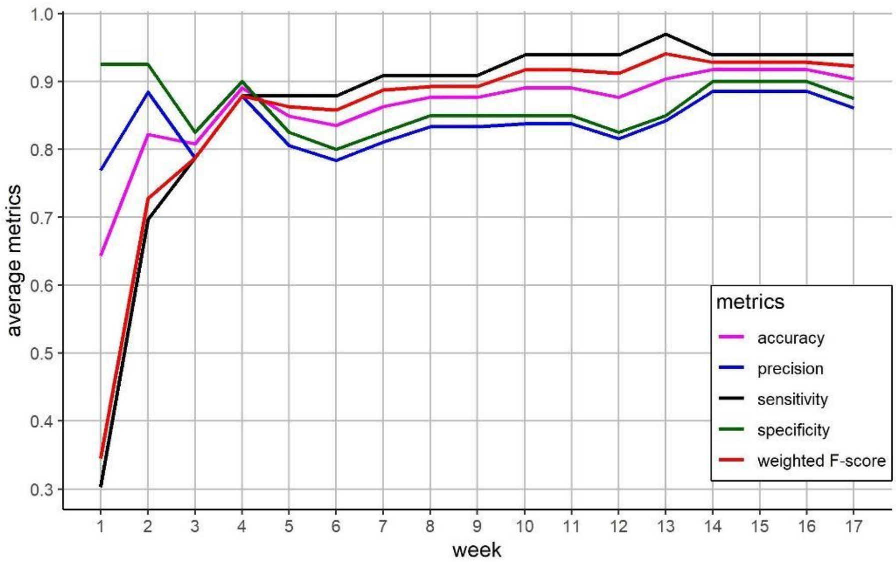

Fig. 5. Classification performance metrics for BNC model on new data

This improvement is most likely due to the fact that the training set for the model was widened by adding new observations, so that the prior distributions in the Bayesian Network were computed more exactly. Sensitivity and the weighted F-score now exceed 86% starting from week 4, which is valuable from the viewpoint of the possibility of earlier detection of at-risk students and their timely support. Starting from week 10, sensitivity exceeds 93% together with the weighted F-score exceeding 91%. The other metrics take values from 78% to 92% for the period starting from week 4 and till the end of a semester.

This result confirms that the BNC model is capable of producing a reasonably accurate prediction from the early stages of the semester.

Simultaneously with testing the model on new data we implemented a student feedback mechanism which informed student about being-at risk of not completing the course. All students engaged in the course were divided into two groups, a control group consisting of 38 students, and an experimental group consisting of 37 students. Starting from the very beginning of the semester we made the forecast weekly and send messages in the LMS to at-risk student from the experimental group. The messages contained information about the level of risk, links to the course elements that should have been completed by that moment and some recommendations about improving the learning performance.

In [36] we analyze the effectiveness of such feedback and find that students from the experimental group are at average: 1) better at attending practical classes; 2) more active when working in the electronic course; 3) more successful in the final exam. However, these results cannot be regarded as statistically significant (since p-values of the Kruskal-Wallis test for the corresponding three hypotheses are 0.2, 0.3 and 0.1). Thus more research is needed on the positive effect of the feedback mechanism on learning success.

By now we have introduced the initial version of EWS in the course of probability and statistics, which includes the BNC model and the described above feedback mechanism aimed at informing students about being at risk and providing them with recommendations on how to improve the situation.

5. Discussions

It is important to point out the benefits of the approach being introduced to timely detection of at-risk students as well as possible difficulties with its implementation.

High quality of forecasts is the obvious goal that all developers of predictive models strive to achieve. However, the usefulness of other principles depends on the learning environment and the stakeholders of the learning process.

For instance, when analyzing large amounts of data in massive online courses, interpretability might not be necessary. Hence, the use of non-interpretable black-box models which often cope better with this kind of data is reasonable. One can find thorough descriptions of such models in the works partially mentioned in the literature review (Section 2).

Unlike the above situation, educational institutions have a wide range of tools to influence the educational process and students' learning behavior. In this case, the value of interpretability increases significantly, because interpretable predictive models not only can solve the main problem of identifying underperforming students, but also contribute to the effective management of the educational process.

Another principle used – adaptability of a predictive model to a course – is also beneficial mainly for courses taught in a blended-learning format. Moreover, we regard it as a crucial property of a predictive model as it enables educators to use the results of in-class learning, which significantly varies depending of the subject area and at the same time makes a great contribution to the success of the students' mastering of the curriculum.

During the pilot study, we were able to identify possible challenges in the process of implementation of the introduced approach.

While *interpretability* and good *overall predictive accuracy* are primarily ensured by the right choice of learning failure predictors and classifiers, compliance with other principles requires some additional work on a learning course.

We list the requirements to a course design, which are essential for *reliable predictions at early* stages as well as for *actionability*:

● There should be a variety of assessment tools for *continuous* measurement of student performance and monitoring of student learning behavior

● It is especially important that *the first half of a semester* is sufficiently equipped with such tools to achieve high quality of predicting learning failure at early stages.

The desired *adaptability* of a predictive model to a course design might face the following challenges:

● The model developers and the teaching staff should work in a close cooperation for better reflecting the course features in the model. Here we have a risk to face teacher resistance at least due to their additional workload or unwillingness to change.

● For a blended-learning format, not all student data is collected by LMS, and there exist several sources of raw data about in-class learning with relatively limited opportunities for managing data extraction. Consequently, it is essential to develop a Single Window system to facilitate the process of data collection and processing.

Our pilot study resulted in the predictive models designed for a certain education course according to this approach. In our case, among other models we chose the Bayesian Network Classifier as the most accurate in the early stages. The design of the model made it possible to evaluate the effectiveness of various assessment tools in at-risk students' detection. For example, it turned out that in the first half of the semester, attendance records are good predictors of academic success, while in the second half of the semester, the importance of this predictor noticeably decreases.

We designed and implemented into the course a partially automated student warning system based on this classifier. Nevertheless, we do not limit ourselves in a range of predictive models while applying the approach to other courses. It is most likely that the resulting early warning system (at the university level) will contain various predictive models and their ensembles.

6. Conclusion

The first objective of the current study is to develop an approach to timely student-at-risk detection, which could serve as a basis for the development of an early warning system. We propose basic requirements for a predictive model, namely: *accuracy*, *interpretability*, *actionability*, *adaptability to a course design* and *high predictive performance at early stages*.

Following the bottom-up approach, we conducted a pilot study for a particular course. We took three classifiers, satisfying the mentioned requirements, and examined their performance through an empirical study. As a result, for the educational course on Probability and Statistics, we have chosen the Bayesian Network Classifier as a most promising predictive model and implemented into the course a partially automated student warning system based on this classifier.

The results of our study show that the proposed approach to development of a predictive model can be valuable in terms of the educational process management. Apart from good at-risk student detection, it provides stakeholders with some additional information on the learning process, such as causal relationships between learning performance in different modules, strengths and weaknesses of a course design, etc.

The development of an early warning system based on such predictive models can provide both teaching and administrative staff with better opportunities of tracking student performance in real-time and timely support students with high enough risk of dropouts. The suggested methodology can be used for designing, developing and implementing an early-warning system at a higher education institution.

***Acknowledgments*:** This work is supported by the Russian Science Foundation under Grant No 22-28-00413.

References

1. F e r g u s o n, R. Learning Analytics: Drivers, Developments and Challenges. – International Journal of Technology Enhanced Learning, Vol. **4,** 2012, No 2, pp. 304-317.
2. G r e l l e r, W., H. D r a c h s l e r. Translating Learning into Numbers: A Generic Framework for Learning Analytics. – Journal of Educational Technology & Society, Vol **15**, 2012, No 3, pp. 42-57.
3. K l e i n, C., R. M. H e s s. Using Learning Analytics to Improve Student Learning Outcomes Assessment: Benefits, Constraints, & Possibilities. – In: Learning Analytics in Higher Education. Routledge, 2018, pp. 140-159.
4. S i e m e n s, G., D. G a s e v i c. Guest Editorial-Learning and Knowledge Analytics. – Journal of Educational Technology & Society, Vol. **15**, 2012, No 3, pp. 1-2.
5. A v e l l a, J. T., M. K e b r i t c h i, S. G. N u n n, T. K a n a i. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. – Online Learning, Vol. **20**, 2016, No 2, pp. 13-29.
6. I f e n t h a l e r, D., D. K. M a h, J. Y.-K. Y a u. Utilising Learning Analytics for Study Success: Reflections on Current Empirical Findings. – In: D. Ifenthaler, D. K. Mah, J. K. Yau, Eds. Utilizing Learning Analytics to Support Study Success. Springer, 2019, pp. 27-36.
7. P a u r a, L., I. A r h i p o v a. Cause Analysis of Students' Dropout Rate in Higher Education Study Program. – Procedia – Social and Behavioral Sciences, Vol. **109**, 2014, pp.1282-1286.

8. B a n e r e s, D., M. E. R o d r í g u e z-G o n z a l e z, M. S e r r a. An Early Feedback Prediction System for Learners At-Risk within a First-Year Higher Education Course. – IEEE Transactions on Learning Technologies, Vol. **12**, 2019, No 2, pp. 249-263.
9. K u h, G. D., J. L. K i n z i e, J. A. B u c k l e y, B. K. B r i d g e s, J. C. H a y e k. What Matters to Student Success: A Review of the Literature. Washington, DC: National Postsecondary Education Cooperative, 2006.
10. F r a z e l l e, S., A. N a g e l. A Practitioner's Guide to Implementing Early Warning Systems. REL 2015-056. Regional Educational Laboratory Northwest, 2015.
11. D h a w a n, S. Online Learning: A Panacea in the Time of COVID-19 Crisis. – Journal of Educational Technology Systems, Vol. **49**, 2020, No 1, pp. 5-22.
12. M ü l l e r, A. M., C. G o h, L. Z. L i m, X. G a o. COVID-19 Emergency e-Learning and Beyond: Experiences and Perspectives of University Educators. – Education Sciences, Vol. **11**, 2021, No 1, p. 19.
13. B a b e r, H. Determinants of Students' Perceived Learning Outcome and Satisfaction in Online Learning during the Pandemic of COVID-19. – Journal of Education and e-Learning Research, Vol. **7**, 2020, No 3, pp. 285-292.
14. Y o r k, T. T., C. G i b s o n, S. R a n k i n. Defining and Measuring Academic Success. – Practical Assessment, Research, and Evaluation, Vol. **20**, 2015, No 1, p. 5.
15. H o m m e l, M., A. E g e t e n m e i e r, U. M a i e r. Supporting Stakeholders with Learning Analytics to Increase Study Success. – In: D. Ifenthaler, D. K. Mah, J. K. Yau, Eds. Utilizing Learning Analytics to Support Study Success. Springer, 2019, pp. 37-60.
16. W o n g, J., M. B a a r s, B. B. d e K o n i n g, T. van der Z e e, D. D a v i s, M. K h a l i l, G.-J. H o u b e n, F. P a a s. Educational Theories and Learning Analytics: From Data to Knowledge. – In: D. Ifenthaler, D. K. Mah, J. K. Yau, Eds. Utilizing Learning Analytics to Support Study Success, Springer 2019, pp. 3-25.
17. O r n e l a s, F., C. A. O r d o n e z. A Naive Bayesian Application to Community College Data. – Technology, Knowledge and Learning, Vol. **22**, 2017, No 3, pp. 299-315.
18. B e k e l e, R., W. M e n z e l. A Bayesian Approach to Predict Performance of a Student (BAPPS): A Case with Ethiopian Students. – Algorithms, Vol. **22**, 2005, No 23, p. 24.
19. M a c f a d y e n, L. P., S. D a w s o n. Mining LMS Data to Develop an "Early Warning System" for Educators: A Proof of Concept. – Computers & Education, Vol. **54**, 2010, No 2, pp. 588-599.
20. B y s t r o v a, T., V. L a r i o n o v a, E. S i n i t s y n, A. T o l m a c h e v. Learning Analytics in Massive Open Online Courses as a Tool for Predicting Learner Performance. – Educational Studies, 2018, No 4, pp. 139-166.
21. O z e r o v a, G. P., G. F. P a v l e n k o. Prediction of Student Performance in Blended Learning Utilizing Learning Analytics Data (in Russian). – Novosibirsk State Pedagogical University Bulletin, Vol. **9**, 2019, No 6, pp. 73-87.
22. K o t s i a n t i s, S., N. T s e l i o s, A. F i l i p p i d i, V. K o m i s. Using Learning Analytics to Identify Successful Learners in a Blended Learning Course. – International Journal of Technology Enhanced Learning, Vol. **5**, 2013, No 2, pp. 133-150.
23. L e e, S., J. Y. C h u n g. The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. – Applied Sciences, Vol. **9**, 2019, No 15, p. 3093.
24. W a n g, W., H. Y u, C. M i a o. Deep Model for Dropout Prediction in MOOCs. – In: Proc. of 2nd International Conference on Crowd Science and Engineering, 2017, pp. 26-32.
25. C h e n, F., Y. C u i. Utilizing Student Time Series Behaviour in Learning Management Systems for Early Prediction of Course Performance. – Journal of Learning Analytics, Vol. **7**, 2020, No 2, pp. 1-17.
26. S o u s a-V i e i r a, M. E., J. C. L ó p e z-A r d a o, M. F e r n á n d e z-V e i ga, O. F e r r e i r a-P i r e s, M. R o d r í g u e z-P é r e z, R. F. R o d r í g u e z-R u b i o. Prediction of Learning Success/Failure via Pace of Events in a Social Learning Network Platform. – Computer Applications in Engineering Education, Vol. **26**, 2018, No 6, pp. 2047-2057.
27. A k c a p i n a r, G., A. A l t u n, P. A s k a r. Using Learning Analytics to Develop Early-Warning System for At-Risk Students. – International Journal of Educational Technology in Higher Education, Vol. **16**, 2019, No 1, p. 40.

28. P o p c h e v, I. P., D. A. O r o z o v a. Towards a Multistep Method for Assessment in e-Learning of Emerging Technologies. – Cybernetics and Information Technologies, Vol. **20**, 2020, No 3, pp. 116-129.
29. K y t m a n o v, A. A., M. V. Noskov, K. V. S a f o n o v, M. V. S a v e l y e v a, V. A. S h e r s h n e v a. Competency-Based Learning in Higher Mathematics Education as a Cluster of Efficient Approaches. – Bolema: Mathematics Education Bulletin, Vol. **30**, 2016, No 56, pp. 1113-1126.
30. N o s k o v, M. V., M. V. S o m o v a, I. M. F e d o t o v a. Management of the Success of the Student's Learning Based on the Markov Model (in Russian). – Informatics and Education, 2018, No 10, pp. 4-11.
31. L i u, R., K. R. K o e d i n g e r. Going Beyond Better Data Prediction to Create Explanatory Models of Educational Data. The Handbook of Learning Analytics, 2017, pp. 69-76.
32. R o s é, C. P., E. A. M c L a u g h l i n, R. L i u, K. R. K o e d i n g e r. Explanatory Learner Models: Why Machine Learning (Alone) is not the Answer. – British Journal of Educational Technology, Vol **50**, 2019, No 6, pp. 2943-2958.
33. G a r d n e r, J., C. B r o o k s. Student Success Prediction in MOOCs. – User Modeling and User-Adapted Interaction, Vol. **28**, 2018, No 2, pp. 127-203.
34. K o r b, K. B., A. E. N i c h o l s o n. The Causal Interpretation of Bayesian Networks. – In: Innovations in Bayesian Networks, Springer, Berlin, Heidelberg, 2008, pp. 83-116.
35. K u s t i t s k a y a, T. A., A. A. K y t m a n o v, M. V. N o s k o v. Student-at-Risk Detection by Current Learning Performance Indicators Using Bayesian Networks. – arXiv preprint arXiv:2004.09774, 2020.
36. K u s t i t s k a y a, T. A., O. A. K a r n a u k h o v a. Developing an Early Warning System to Detect At-Risk Students: A Feedback Mechanism. – In: Proc. of 5th International Conference on Informatization of Education and e-Learning Methodology: Digital Technologies in Education, 2020, pp. 289-293.

Received: 02.12.2021; Second Version: 04.02.2022; Accepted: 21.02.2022