MATHEMATICAL MODELS OF LEARNING ANALYTICS FOR
MASSIVE OPEN ONLINE COURSES

E. Sinitsyn, A. Tolmachev, V. Larionova, A. Ovchinnikov
Ural Federal University (RUSSIAN FEDERATION)

Abstract
Online education is rapidly developing in Russia and in the world over the last decades. One of the most popular and available online learning technologies is massive open online courses (MOOC), which are successfully used in the university, school, continuing professional education and informal life-long learning. Nowadays the number of MOOC learners in the world is estimated in tens of millions of people. Nevertheless, using MOOC in the educational process has both advantages and disadvantages. The latter include problems connected with individualization of training, assessment of progress and support for students, and assessment of the quality of courses. Analyzing and predicting the success of students is an important task and tool for solving them. The paper is devoted to developing a model of students’ progress forecasting, which is based on the information theory and allows one to estimate the probability of students’ success in the of the final test when observing the current performance of students. The probabilistic model for group forecasting of the final performance, considered in the paper, allows predicting the distribution of students’ final scores and reduce the level of uncertainty in online learning using MOOCs. It can be useful for assessing the quality of the course as a whole and developing measures to improve it. The statistical model for personalized performance prediction allows one to make a forecast of the final progress for each student and can be used to identify negative trends and problems of students in learning, provide them with relevant feedback and necessary support. The application of the models is considered on the example of an engineering online course created by the Ural Federal University

Keywords: MOOC, learning analytics, probabilistic and statistical methods, information theory, learning performance forecasting.

1 INTRODUCTION
Online learning has been rapidly developing in Russia and in the world over the last decades. It is used in the educational process of higher educational institutions, corporate universities, schools and organizations implementing professional training programs. One of the most popular and available online learning technologies is massive open online courses (MOOC) [1]. According to Holon IQ [2] nowadays MOOCs have over 80 million learners in the world. By 2030, the number of graduates from schools and universities will increase by about a billion people compared with 2015, which will significantly increase the demand for both traditional and online forms of education.

In the global open education market, the leaders among MOOCs platforms are Coursera, Udacity, edX, Udemy, and others. They offer more than ten thousand online courses including: online courses from the best universities in the world; professional courses for additional training, which are recognized by the world's largest corporations; corporate courses for employees of partner companies; general courses for life-long learning [3]. They can be a part of full bachelor or master degree programs on some specific specialisms or single online programs, which can last from a few months to 3–7 hours (mini-format) or 2 hours (micro-format).

An example of Russian open education platforms can be viewed on the Open Education Platform (https://openedu.ru/) [4]. To date, 353 online courses from 14 leading Russian universities, including Moscow State University, Moscow Institute of Physics and Technology, Ural Federal University, Higher School of Economics, and others are presented on the platform [5]. Students of these universities, other Russian universities, and anyone wishing to study can the take courses and obtain credits for them [6].

Online learning has both advantages and disadvantages. The advantages include: financial affordability, which is expressed in lower costs as compared with the traditional education; availability of content, which gives the opportunity for distance study form any place at any time; adaptability, which enables students to choose courses from the best teachers, as well as the pace and form of learning; the presence of a large amount of data on the activity of students during the course,
available for analysis. The disadvantages are due to the specific features of the MOOC - a large number of students, and the lack of personal contact during the learning process. The disadvantages give rise to new challenges for the education system. First of all, how does the provider ensure an individual approach for each student with the necessary contact with a teacher in the online format as the students form individual learning paths. Secondly, how is the progress of students assessed in self-paced training, how is success predicted, how are problems detected in a timely manner, how to provide support for students, how to increase retention and improve the share of students who successfully complete the course. Third, how to assess the quality of online courses, how to determine the relevance of assessment tools and improve those tools.

This paper discusses mathematical models developed to provide answers to the above-listed challenges.

2 METHODOLOGY

When students take an online course, large amounts of data are accumulated. They are:

- big data at the educational platform level, including students’ assessments when they pass current control events and the final test, log files with information about students’ activity when they work with different components of course content (number of communications and attempts, time and duration of various activities, engagement and satisfaction of learners, etc.);
- data from other sources, including information about the performance of students in other courses, results of surveys, personal data from the university or provider administrative base and information about other achievements and activities of students.

These data represent a digital trail of students, which can be used for analysis and prediction of the outcomes of students' training. Support for students can be optimized, adaptive learning organized by constructing individual educational paths, course quality can be assessed and recommendations made for course improvement. Modern technologies of data mining and machine learning are applied in the analysis of the digital trail.

In this paper, we develop mathematical models for predicting students’ potential for achieving the learning outcomes in online courses. The input data of the models are information about students’ current performance, which is generated when students pass control points provided by the course program. The use of such data has several advantages, including:

- data of current progress and learning outcomes are recorded objectively regardless of any model of the educational process;
- academic performance in higher education institutions is often considered as the basis for administrative decisions (for example, decisions on students' dismissal or award of scholarships), therefore early warning and prevention of students' dropouts are important;
- learning analytics allows assessment of the quality of course control tools and their relevance in measuring the learning outcomes.

Further, a learning analytics model of students’ performance analysis, a probabilistic model for group prediction of the final test score distribution, and a statistical model for personal forecasting of students' online course progress will be considered in detail. These outputs represent a digital trail of students, which can be used to analyze and make decisions aimed at predicting the effectiveness of students' training, ensuring their support during the course, organizing adaptive training with building individual educational trajectories, assessing the quality of courses and developing recommendations for their improvement. The digital footprint can be explored deeply through modern technology of machine learning and data mining.

3 RESULTS

3.1 Performance Model Analysis – based on information theory

Consider a MOOC, which provides for N current checkpoints, as well as a final test. Current checkpoint forms a system of N elements:

\[ X = \{X_1, X_2, ..., X_N\} \] (1)
In addition, each element $X_i$ is a vector of grades (the number of grades received) of all students according to the results of the $i$-th checkpoint. The grades received by the course participants for the final test will be denoted by the vector $Y$.

In the practice of Russian universities, a 100-point system is usually used, but in this work in order to analyze and predict it is more convenient to use quantized values of grades for $X$ and $Y$ with a larger step:

Table 1. A system of grades for the current assessment event and the final test.

<table>
<thead>
<tr>
<th>$k$</th>
<th>Grades</th>
<th>Range of points received</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NA</td>
<td>The student did not pass the control event</td>
</tr>
<tr>
<td>2</td>
<td>Failure</td>
<td>Less than 40 points</td>
</tr>
<tr>
<td>3</td>
<td>Pass</td>
<td>More than 40, but less than 60 points</td>
</tr>
<tr>
<td>4</td>
<td>Success</td>
<td>More than 60 points</td>
</tr>
</tbody>
</table>

As shown in Table 1, an increase in the number of categories for grades does not make sense. For example, the division of the {Success} group into the {Good} and {Excellent} groups in accordance with the traditions of the Russian education system only complicates the analysis and makes it more cumbersome, without leading to either a qualitative change in the results or an increase in the accuracy of forecasts.

Thus, in the MOOC training process, for each student, data are accumulated in two interrelated systems: the system of current performance checkpoints (1) - $X$, and the associated final testing system - $Y$.

If we approach the analysis of the relationship of these systems based on the information theory, then we can rely on the ability to predict the results of the final test only if the systems $X$ and $Y$ are interdependent, and obtaining information about the system $X$ reduces the level of uncertainty about the system $Y$.

To quantify the amount of information about the system $Y$, which can give an observation of the system $X$, we use the standard formula:

$$I_{X \rightarrow Y} = H(Y) - H(Y/X) = H(X) - H(X/Y)$$

where $H(Y/X)$ and $H(X/Y)$ conditional entropy of interconnected systems $X$ and $Y$ [7]:

$$H(Y) = -\sum_{k=1}^{4} P_k \cdot \log_2(P_k)$$

$$H(Y/X) = -\sum_{k=1}^{4} P_k \cdot P(y_k/x_1,x_2,\ldots,x_N) \cdot \log_2(P(y_k/x_1,x_2,\ldots,x_N) \cdot P_k)$$

Here $P(y_k/x_1,x_2,\ldots,x_N)$ - the conditional probability of the event that a learner will obtain the grade $y_k$ in the final test ($k$ – is presented in Table 1), if he obtained the grade $x_1$ in the test $X_1$, the grade of $x_2$ in the test $X_2$, $x_N$ - in the test $X_N$. $P_k$ – the probability of obtaining the grade $y_k$ in the final test.

The use of the logarithm of base 2 automatically entails the measurement of information and the entropy of systems $X$ and $Y$ in bits.

All the probabilities listed above and the amount of information about the final test that can be obtained from observations of current performance are determined on the basis of data accumulated by the educational platform.

For illustration, consider the course “Engineering Mechanics” for the fall semester of 2017, developed and implemented by the Ural Federal University (Russia). Figure 1 shows the probability of the course participants obtaining grades for the final test in accordance with Table 1. The number of learners considered when calculating probabilities is 918.
Presenting the conditional probability is problematic because of the large number of variables $x$ that the course includes - 37 types of different checkpoints. The results of calculations (2) - (4) for this course are the following:

$$H(Y) = 1.51 \text{ bits} ; \quad H(X) = 6.92 \text{ bits} ; \quad H(Y/X) = 1.01 \text{ bits} ; \quad H(X/Y) = 6.42 \text{ bits} ; \quad I_{X\rightarrow Y} = 0.49 \text{ bits} \quad (5)$$

It is also useful to evaluate the joint entropy of the systems $X$ and $Y$ – $H(X,Y)$, which is a measure of their interdependence:

$$H(X,Y) = 7.93 \text{ bits} \quad (6)$$

Let us compare the entropies of the systems $X$ and $Y$ with the data that would have been obtained with purely random and equal probable rating according to the test results. As noted in [8], on the one hand, this makes it possible to assess the quality of the tests themselves; a low level of uncertainty of the test results indicates that the test is not informative (for example, no one passes it or not), on the other hand, a high level of uncertainty compromises tested and/or testers because it can correspond to a purely random answer and/or rating.

For purely random and equally likely rating:

$$H_{R}(X) = 2 + \log_{2}(N) \text{ bits} ; \quad H_{R}(Y) = 2 \text{ bits} ; \quad H_{R}(Y/X) = H_{R}(Y) ; \quad H_{R}(X/Y) = H_{R}(X) \quad (7)$$

For the considered course $N = 37$, therefore, $H(X) / H_{R}(X) = 96\%$; $H(Y) / H_{R}(Y) = 75.4\%$. Thus, if the uncertainty of the final test lies within reasonable limits, then the estimates of the current performance are close to the result that could be obtained with a random arrangement of the grades.

The results show that, initially, prior to the analysis of the current performance data, the uncertainty of the final test scores is 1.51 bits. Monitoring students' current performance reduces this uncertainty by 0.49 bits, or about 33%. Accordingly, algorithms for predicting the results of final testing based on the analysis of current performance data cannot lead to a greater reduction in uncertainty. Note that this is a lot. For example, the uncertainty (entropy) of a forecasting system that has two possible states: $H_{1} = "the forecast is correct"; \quad H_{2} = "wrong forecast"$; reduced by 30% with a probability of a correct prediction of about 80%.

Let's consider how the uncertainty of the results of the final testing changes as students are passing the course, that is, as you move from the beginning of the course to the end. In Figure 2, this relationship is shown for the course “Engineering Mechanics” at UrFU autumn 2017.
Figure 2. Reducing a degree of uncertainty of the results of the final test as the students are passing checkpoints (from the first to the current with the number $i = 1 \ldots 37$)

It is clearly seen that even monitoring the students’ performance of the first checkpoint reduces the uncertainty of the results of final testing to the level determined by (5).

One can also consider the partial probability information [7], which reduces the uncertainty of the results of final testing and is contained in the information that the checkpoint number $i$ was passed by the learner with one of the grades presented in Table 1. These data are shown in Figure 3.

Figure 3. Partial probability information that reduces the uncertainty of the final test contained in the information about the results of passing checkpoints by the learner.

As can be seen, the most informative message is that the learner did not go through the checkpoint (NA). However, this is of little interest, since it most likely reflects the fact that students who have not completed a sufficiently large number of control tasks, as a rule, do not try to pass the final test. The second and third place on informativeness is occupied by the information that the checkpoint was not satisfactorily completed (Failure) or successfully (Success). The least informative is the information that the checkpoints were passed only satisfactorily (Pass).

Thus, our joint analysis of the results of final testing and current checkpoints shows that information about the latter objectively reduces the level of uncertainty in the results of final testing and forecasting.
final testing scores based on the current checkpoints can be quite successful. In the considered example, the level of reduction in the uncertainty of the final testing (entropy) [7] reached 33%.

### 3.2 Probabilistic Model of Group Forecasting Performance Prediction

In order to construct a mathematical model for group forecasting of academic performance, we reduced the number of groups in table 1 to three, combining groups of students with “Failure” and “NA” grades into one - “Failure”.

Every checkpoint of the course is described by the vector \( X = \{x_1, x_2, x_3\} \), where \( x_i \) is the ratio of the students’ quantity in each of three groups \( X_1, X_2, X_3 \). Passing of checkpoints is accompanied by transitions of students between groups. In this case, the transition probabilities can be represented as a matrix:

\[
\hat{P} = \begin{bmatrix}
P_{11} & P_{12} & P_{13} \\
P_{21} & P_{22} & P_{23} \\
P_{31} & P_{32} & P_{33}
\end{bmatrix}
\]  

Matrix \( \hat{P} \) is asymmetric and for its elements holds:

\[
\sum_{j=1}^{\alpha^2} P_{ij} = 1 
\]  

Using a probabilistic model, we can predict the distribution of students in groups \( X_1, X_2, X_3 \); this prediction is group based and does not contain personal information.

The change in the number of students in groups 1, 2, 3 can be considered as a Markov process associated with transitions between groups at times \( t_1, t_2, ..., t_\zeta \). If \( P(X_t, X_r, X_s|t) \) is the probability that groups 1, 2, 3 will contain \( X_1, X_2, X_3 \) members at the moment of time \( t \), then equation for \( P(X_t, X_r, X_s|t) \) is written as follows [9]:

\[
\frac{\partial P(X_1, X_2, X_3|t)}{\partial t} = P(X_1, X_2, X_3|t) \cdot \left( (1-z) \cdot \sum_{i=1}^{\alpha} P_{ii} - \sum_{i=1}^{\alpha} X_i \right) + z \cdot \sum_{i=1}^{\alpha} (X_i + 1) \cdot P(\ldots, X_{i+1}, \ldots|t) + (1-z) \cdot \sum_{i=1}^{\alpha} \sum_{j, j \neq i} P_{ij} \cdot (X_i + 1) \cdot P(\ldots, X_{i+1}, \ldots X_{j-1}, \ldots|t)
\]

Let us assume \( z = 0 \) (\( z \) is the probability of a student dropout per unit of time) and exclude the corresponding entries from the initial data.

As shown in [8] and [9], in the case of a large number of students \( N \) (which is applicable to the MOOCs), their distribution among groups is not random and the number of members in the \( i \)-th group is close to the average \( \langle X_i \rangle \). Where \( \langle X_i \rangle = \langle X_2 \rangle = \langle X_3 \rangle = N \).

The equation for \( X_i \) can be written as following:

\[
\frac{\partial X_i}{\partial t} = \sum_{k=1}^{\alpha} [\vec{P}_{ki} \cdot X_k - \vec{P}_{ik} \cdot X_i],
\]

where

\[
\vec{P}_{ki} = 0 \text{ for } k = i, \quad \vec{P}_{ki} = P_{ki, defined by (8)} \text{ for } k \neq i
\]

In the stationary case \( \frac{\partial X_i}{\partial t} = 0 \) it is possible to calculate from (5) the elements of the transition matrix (8) and its eigenvectors \( \vec{a}_x = \hat{P} \cdot \vec{a}_x \). It is known from [10] that the transition matrix (8) can be associated with the task of random walks on an oriented graph whose vertices correspond to the groups \( X_1, X_2, X_3 \), and the transition probabilities between the vertices are determined from the transition matrix (2). Vector \( \vec{a}_x = \{x_1, x_2, x_3\} \) can be viewed as the steady-state distribution of students that occurs after multiple transitions \( \vec{a}_x(n) = \hat{P} \cdot \vec{a}_x(n-1) \) and \( n \rightarrow \infty \). This limit case corresponds to the hypothetical situation of repeated passage of the considered checkpoint by groups of students with statistically equivalent characteristics of learning performance.

If the transition matrix \( \hat{P} \) (8) is known, then according to the theory of Markov processes [11], it can be used to predict the results of passing checkpoints:

\[
\vec{a}_x(1) = \hat{P} \cdot \vec{a}_x(0),
\]

where \( \vec{a}_x(0) \) is vector of student distribution before passing the checkpoint, \( \vec{a}_x(1) \) - after passing it.
Using the probabilistic model for the training sample set, the transition matrix (8) was calculated for the course “Engineering Mechanics” (Fall 2017) offered by Ural Federal University. The steady-state distribution for the transition from the average current learning performance for all checkpoints to the final test was as follows:

\[
\bar{\rho} = \begin{bmatrix}
N & P1->N & P2->N & P3->N \\
1 & 0,25 & 0,175 & 0,077 \\
2 & 0,25 & 0,2 & 0,109 \\
3 & 0,5 & 0,625 & 0,814
\end{bmatrix}
\]

\[
\bar{\alpha}_x = \begin{bmatrix}
\text{Category} & \text{Part} & \% \\
X_1 (failure) & 0,109 & 10,9% \\
X_2 (pass) & 0,137 & 13,7% \\
X_3 (success) & 0,754 & 75,4%
\end{bmatrix}
\]

The results of group forecasting for the student distribution after the final test based on the average current learning performance for the entire course, performed according to (13) on the test sample set, are presented in Table 1. Because the difference between actual and predicted distributions by groups is not statistically significant (p > 0.05 in Pearson’s Chi-squared test), they can be considered as equivalent.

Table 2: The results of group forecasting using a probabilistic model.

<table>
<thead>
<tr>
<th>Group</th>
<th>Average current grade, students (%)</th>
<th>Final test (in fact) students (%)</th>
<th>Final test (forecast) students (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_1 (failure)</td>
<td>5 (5,6%)</td>
<td>5 (5,6%)</td>
<td>9 (10,1%)</td>
</tr>
<tr>
<td>X_2 (pass)</td>
<td>11 (12,4%)</td>
<td>15 (16,9%)</td>
<td>11 (12,4%)</td>
</tr>
<tr>
<td>X_3 (success)</td>
<td>73 (82,0%)</td>
<td>69 (77,5%)</td>
<td>69 (77,5%)</td>
</tr>
<tr>
<td>Total</td>
<td>89 (100%)</td>
<td>89 (100%)</td>
<td>89 (100%)</td>
</tr>
</tbody>
</table>

The representation of the transition matrix in the form of an oriented graph whose vertices correspond to the groups X_1, X_2, X_3, and the transitions correspond to random walks with probabilities (8), is shown in Figure 4.

![Figure 4. Oriented graph for transitions between X_1, X_2, X_3 (1) for the final test with probabilities (8).](image-url)
3.3 The statistical model for personal performance forecasting

The statistical model proposed by the authors is based on traditional methods of data analysis and data science [12] - clustering and regression. The model provides a personal forecast for the passing success of the final test on the basis of current learning performance. The calculations involve students who are active during the course and have an average score for the current learning performance greater than zero. The algorithm of the statistical model for a random checkpoint (C) follows:

- Randomly select two sample sets of students - training and test. Calculate for each student from the training and test sample sets the average current learning performance by averaging the grades from the first checkpoint to the current checkpoint (C).

- Clusterize the training sample set by the average current learning performance using the k-means method. Calculate the linear regression parameters for each cluster of the training sample set. The independent variable is the average current learning performance. The dependent variable is the grade for the final test.

- Clusterize the test sample set by the average current learning performance using the k-means method. Apply linear regression with the parameters according to previous step to the clusters of the test sample sets and calculate predicted values of grades for the final test. Classify the forecast and actual grades for the final test according to the table, calculate the number of correct and incorrect forecasts, evaluate the statistical significance of the results.

In case of deterioration in the learning performance forecast for any student, the course administrator can timely support and motivate the student to improve the indicators.

Using statistical modelling, prediction of the personal results for the final test was performed for the course “Engineering Mechanics” (Fall 2017) offered by Ural Federal University. Calculations on the basis of the average current learning performance were made for all 37 checkpoints. The formation of training and test sample sets was carried out randomly from the initial set of students in a ratio of 80/20 respectively. Three variants of the number of clusters were considered: 1, 2, 3.

The forecasting results are presented in Figure 5. The percentage of correct forecasts is about 70%. Visually, the best result comes from using two clusters. However, an analysis of the results shows that they are not statistically significantly different (p > 0.05 in Pearson’s Chi-squared test), therefore clustering does not provide a significant improvement in predictions for this course.

![Figure 5. The percentage of the correct forecasts for a different number of clusters in case of statistical model.](image)
The reasons for about 30% of incorrect forecasts lie outside the statistical model. Let us consider the contingency tables "Final test (forecast)" - "Average current grades" and "Final test (actual)" - "Average current grades" after 39 checkpoints for two clusters, which are shown in Tables 3 and 4 respectively.

Table 3. Contingency table "Final test (forecast)" - "Average current grades"

<table>
<thead>
<tr>
<th>Average current grades</th>
<th>0.0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td>3</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
<td>3</td>
<td>18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
<td>20</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>18</td>
<td>1</td>
<td>10</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Contingency table "Final test (actual)" - "Average current grades"

<table>
<thead>
<tr>
<th>Average current grades</th>
<th>0.0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td>14</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

We can see that the predictions using the statistical model based on clustering and regression lie on the diagonal of the contingency table. It means that the better average current learning performance corresponds to the better result of the final test. In this case, deviations of elements from the table diagonal will produce incorrect decisions. For example, the elements highlighted in the Table 4 (~18% of the test sample set) represent students with average current grade ≥ 0.4 and with 0.0 for the final test (case of incorrect forecasts). At the same time there are elements lying on the diagonal of the Table 4 and represent students with average current grade and grade for the final test ≥ 0.4 (case of correct forecasts). In the first case abnormal student behavior is detected and its reason may be in passing of final test at another university (this fact is not reflected in learning analytics). The statistical model can be improved by additional parameters such as results of the questionnaire which contain information where students will pass the final test.

4 CONCLUSIONS

The study of performance using a model based on information theory showed that monitoring the current performance of students leads to a decrease of the level of uncertainty in the results of final
testing. Using the example of the online course developed by UrFU considered in the paper - by 33% - the results of group and personal performance predictions based on probabilistic and statistical models are significant. The level of correct personal forecasts for the online course of UrFU was about 70%.

These results are important for automating of individual work with students, for operative reactions and support of students, for moving towards adaptive learning, for improvement of materials and course methods. The ways of further improvement of forecasting models and learning analytics data are outlined.

REFERENCES


