



© Г. П. Озерова, Г. Ф. Павленко

DOI: [10.15293/2658-6762.1906.05](https://doi.org/10.15293/2658-6762.1906.05)

УДК 378+004

## Прогнозирование успешности студентов при смешанном обучении с использованием данных учебной аналитики

Г. П. Озерова, Г. Ф. Павленко (Владивосток, Россия)

**Проблема и цель.** Статья посвящена проблемам прогнозирования успешности студентов, обучающихся с использованием *online*-платформ. Цель статьи – выявить особенности прогнозирования успешности студентов при смешанном обучении на основе данных учебной аналитики.

**Методология.** Основными методами исследования являются: теоретический анализ и обобщение научно-исследовательских работ, теоретические и практические методы педагогического исследования, методы статистической обработки эмпирических данных, методы машинного обучения и методы моделирования случайных событий.

**Результаты.** Проведенное исследование выявило, что прогнозирование должно осуществляться на основе критериев, определяющих успешность обучения, метрики для которых можно получить на основе данных учебной аналитики. Классификацию студентов на группы успешности по выбранным критериям необходимо проводить для каждого контрольно-измерительного инструмента непосредственно после его выполнения студентами, чтобы своевременно выявить обучающихся, нуждающихся в особом внимании со стороны преподавателя. Для прогнозирования успешности обучения других потоков студентов целесообразно накапливать информацию о динамике переходов обучающихся между группами успешности, используя дискретные цепи Маркова.

**Заключение.** Прогнозирование успешности студентов на основе данных учебной аналитики позволяет выделить обучающихся «группы риска», предсказывать распределение студентов по группам успешности и при необходимости корректировать учебно-методические материалы.

**Ключевые слова:** успешность обучения; смешанное обучение; система управления обучением; учебная аналитика; прогнозирование; классификация; дискретные цепи Маркова.

### Постановка проблемы

В последние годы смешанное обучение (комбинация традиционного и *online* обучения)

стало одной из самых востребованных образовательных технологий в связи с растущим внедрением систем управления обучением

**Озерова Галина Павловна** – кандидат технических наук, доцент кафедры механики и математического моделирования, Инженерная школа, Дальневосточный федеральный университет.

E-mail: [ozerovalp@dvfu.ru](mailto:ozerovalp@dvfu.ru)

**Павленко Галина Федоровна** – кандидат технических наук, доцент кафедры бизнес-информатики и экономика-математических методов, Школа экономики и менеджмента, Дальневосточный федеральный университет.

E-mail: [pavlenkogf@dvfu.ru](mailto:pavlenkogf@dvfu.ru)

(LMS) в образовательный процесс [1, с. 4]. В России самой популярной системой с открытым и сходным кодом является Moodle, ее используют более 65 % вузов. В ряде университетов внедрена Blackboard, самая распространенная в мире система с закрытым исходным кодом. LMS поддерживает вовлечение студентов в учебный процесс, создает условия для активного взаимодействия студентов и преподавателей, осуществляет сбор данных о поведении обучающихся на online платформе [2, с. 3–4]. Наличие системы организации обучения, достижения в области статистики, быстрое развитие программных и аналитических методов, а также возможность применения принципов бизнес-аналитики к процессу обучения, привели к появлению учебной аналитики как одного из наиболее перспективных направлений исследований в области компьютерной поддержки образования [3, с. 298].

Учебная аналитика – это измерение, сбор, анализ и представление данных об обучающихся и их действиях с целью понимания и оптимизации учебного процесса и той среды, где это этот процесс происходит [4, с. 336–337]. Анализ таких данных позволяет контролировать регулярность занятий слушателя, осуществлять мониторинг его успеваемости, следить за ходом выполнения контрольных заданий [5, с. 145]. В последние годы учебная аналитика послужила концептуальной основой для анализа образовательных данных online курсов [6], прогнозирования успешности обучающихся [9–25], сбора данных об их поведении и своевременного вмешательства преподавателя в процесс обучения [7].

В данной работе проверяется гипотеза о том, что при смешанной модели данные учебной аналитики LMS наряду с экспертным мнением преподавателя позволяют прогнозировать успешность студентов. Понятие «успеш-

ность обучения» имеет разную трактовку, некоторые ученые связывают его с успеваемостью студентов, Б. Г. Ананьев определяет его как «оптимальное сочетание темпа, напряженности, индивидуального своеобразия (стиля) учебной работы, степени прилежания и усилий, которые прилагает обучаемый для достижения определенных результатов» [8, с. 442].

На основе данных учебной аналитики LMS можно получить большой массив эмпирических данных о поведении студента на online платформе, определяющих вовлеченность, непрерывность, результативность, самостоятельность и другие аспекты учебной деятельности студента [9, с. 9]. Модели прогнозирования только на основе этих данных позволяют соотнести текущую деятельность обучающегося с его будущими перспективами, например, «забросить учебу» или получить высокий итоговый результат. В [10, с. 41] отмечается, что такую диагностику необходимо проводить как можно раньше, чтобы у преподавателей было достаточно времени для проведения образовательных мероприятий, способствующих повышению успешности обучения студента.

Одной из возможностей учебной аналитики является накопление данных о поведении студентов, обучавшихся в предыдущие временные периоды. Это позволяет выделить группы студентов со сходным поведением и при прогнозировании учитывать накопленный «опыт», а не только текущую информацию о деятельности студента [11].

В работе [12] отмечается, что для прогнозирования чаще всего используются линейные модели, такие как линейная регрессия [13–15] и логистическая регрессия [16] из-за их простоты и возможности интерпретировать линейные закономерности между данными учебной аналитики и достижениями студентов. В [13] авторы использовали online данные

о поведении более 220 студентов инженерных специальностей для изучения факторов, влияющих на успешность их обучения. В [14] описывается методика прогнозирования прогресса студентов средствами регрессионного анализа на основе более 20 эмпирических показателей. В работе [15] на основе k-medoids кластеризации студенты были разделены на группы успешности выполнения самостоятельной работы. Затем на основе регрессионных моделей сформулированы стратегии поддержки студентов каждой группы.

Метод опорных векторов (Supported Vector Machine), относящийся к алгоритмам линейного машинного обучения с учителем, в контексте прогнозирования использовался для нескольких целей [17; 18]. Например, в [17] авторы прогнозируют отсев обучающихся на массовых открытых онлайн курсах платформ Coursera и EdX. В [18] метод применяется для прогнозирования готовности студента выполнить контрольное задание.

Другим методом прогнозирования является использование деревьев решений (Decision Trees). В [19] с помощью интеллектуального анализа данных и деревьев решений выделяется несколько моделей поведения студентов при online обучении, которые определяют способ взаимодействия обучающегося с содержимым, его длительность и перспективу завершения online курса. В [20] предлагается классификационная модель для выявления студентов, которые могут «забросить» учебу в течение одной недели. Для моделирования используется байесовский классификатор и алгоритм C 4.5 построения дерева решений. Результаты исследования показали, что коэффициент удержания был бы выше, если бы преподаватели своевременно вмешивались в процесс обучения на основе результатов прогнозирования.

В работах [21] для прогнозирования использовался алгоритм «случайный лес» (Random Forest), суть которого заключается в использовании большого количества деревьев решения, каждое из которых само по себе дает очень невысокое качество классификации, но за счет их большого количества получается хороший результат [12]. Пример использования алгоритма для прогноза отсева обучающихся показан в работе [21]. Авторы разработали разные модели деревьев решений, в том числе Random Forest, оценили точность прогноза и показали, что использование последней улучшает точность прогноза до 91 %. Также были отмечены проблемы с производительностью и масштабированием модели.

В работе [5] прогнозирование выполняется на основе вероятностной марковской модели. Авторами предлагается отслеживать индивидуальные траектории прохождения слушателями курса и на их основе определять вероятности переходов между группами слушателей по успеваемости в процессе обучения. Это позволяет предсказывать результаты выполнения заданий курса разными категориями обучающихся.

Кроме перечисленных методов для прогнозирования успешности обучения на основе данных учебной аналитики используются нейронные сети [22], общие байесовские сети [23], модели Probabilistic Soft Logic [24], классификационные модели машинного обучения [25].

Определенная часть элементов учебного процесса при смешанном обучении может быть полностью реализована на online платформе. Прежде всего, это самостоятельная работа студентов, а также выполнение различных контрольно-измерительных мероприятий текущего и промежуточного контроля. Для оценки качества самостоятельной работы применяется совокупность показателей [26,

с. 141–142], которые при смешанном обучении могут быть рассчитаны на основе данных учебной аналитики. Учебную аналитику используют и для анализа выполнения студентами контрольно-измерительных заданий [27, с. 566]. Кроме того, при смешанном обучении преподаватель может непосредственно наблюдать за студентами на аудиторном занятии, его экспертные оценки могут дополнить данные для прогнозирования успешности обучения, что позволит улучшить прогнозирующие характеристики моделей.

*Целью* данной работы является выявление особенностей прогнозирования успешности обучения студентов для дисциплин, реализуемых по технологии смешанного обучения.

#### Методология исследования

Основными методами исследования являются: теоретический анализ и обобщение научно-исследовательских работ, теоретические и практические методы педагогического исследования, методы статистической обработки эмпирических данных, методы машинного обучения и методы моделирования случайных событий.

*Эмпирическая база.* В исследовании использовались данные учебной аналитики о действиях в LMS BlackBoard 538 студентов первого курса Инженерной школы Дальневосточного федерального университета при изучении дисциплины «Информационные технологии». Дисциплина реализовывалась в формате смешанного обучения в 2018–2019 учебном году. Полностью на online платформе студенты выполняли самостоятельную работу и контрольные задания текущей и промежуточной аттестации. Работа преподавателей была организована таким образом, что учет остальных видов деятельности и процесс взаимодействия со студентами по проверке заданий и

проведения консультаций также осуществлялся на online платформе. Всего на online платформе было реализовано 50 контрольных точек (элементов курса, которые предусматривают взаимодействие с обучающимся), объединенных в 5 модулей.

*Переменные исследования.* Независимые переменные, используемые в моделях прогнозирования, и их описание приведено в таблице 1. На выполнение каждого контрольно-измерительного задания студенту отводилось некоторое время, вычисление показателей осуществлялось непосредственно после окончания срока (для раннего прогнозирования) и по окончании изучения модуля (для прогнозирования успешности следующих потоков обучающихся).

*Описание модели для раннего прогнозирования.* Для раннего прогнозирования успешности обучения проводилась классификация студентов на группы методом машинного обучения Random Forest. Количество групп может быть любым, в данном исследовании использовались три группы:

- 1 группа – высокий уровень успешности;
- 2 группа – средний уровень успешности;
- 3 группа – низкий уровень успешности.

Классификация методом Random Forest основана на задаче классификации с учителем. Формально эту задачу можно описать в следующем виде. Пусть  $X$  – множество описаний объектов,  $Y$  – конечное множество номеров классов. Существует отображение  $f: X \rightarrow Y$ , значения которого известны только на объектах конечной обучающей выборки  $X_m = (x_1, y_1), \dots, (x_m, y_m)$ ,  $y_i$  – номера класса. Требуется построить алгоритм  $f_a: X \rightarrow Y$ , способный классифицировать произвольный объект  $x \in X$ .

Таблица 1

## Описание независимых переменных

Table 1

## Predictor Variables

Переменная	Описание
$f_1$ , (дни)	своевременность выполнения: значение показателя равно 0, если задание выполнено в определенный преподавателем интервал времени, в противном случае – количество дней от даты окончания периода и до первой выполненной попытки
$f_2$ , (%)	результативность, определяется как максимальный результат всех попыток в определенный интервал времени, если задание выполнено вовремя, или средний результат попыток, в противном случае
$f_3$ , (час)	периодичность и непрерывность выполнения заданий, определяется как время, проведенное студентом на курсе в заданный период времени
$f_4$ , (%)	самостоятельность выполнения, вычисляется как отношение количества результативных попыток к общему количеству попыток (если студент выполняет слишком много нерезультативных попыток, скорее всего он использует недобросовестные методы выполнения: ищет известные задания и пользуется готовыми ответами)
$f_5$ , (балл от 1 до 10)	результат собеседования и ответов на вопросы, используется только для тех заданий, которые проверяются преподавателем, для заданий, проверяемых автоматически, этот показатель одинаков для всех студентов

В нашем случае объект представляет собой признаковый вектор  $x = (f_1(x), f_2(x), \dots, f_n(x))$ , каждый элемент которого является числовым значением и описывает некоторый аспект поведения студента при выполнении контрольно-измерительного задания. А обучающая выборка строится на основе сформулированных преподавателем набором правил для каждой независимой переменной. Перед применением классификационного алгоритма все входные данные нормировались.

*Описание модели прогнозирования для следующих групп обучающихся.* После выполнения каждого контрольно-измерительного инструмента студенты могут переходить из одной из одной группы успешности освоения дисциплины в другую. Отслеживая такие переходы по каждому студенту, можно рассчитать вероятности переходов между группами

по каждому контрольно-измерительному элементу. Это позволит применить классическую модель Маркова для прогнозирования успешности студентов, проходящих обучение в следующий период времени.

Пусть дисциплина содержит  $n$  контрольно-измерительных инструментов. Для анализа успешности выполнения  $r$ -го контрольно-измерительного инструмента ( $r = 2..n$ ) строится ориентированный граф, вершины которого соответствуют номерам групп, а дуги – количествам переходов студентов из одной группы в другую после выполнения задания  $r$  по отношению к распределению студентов после выполнения задания  $r-1$ . Матрица смежности графа для  $r$ -го контрольно-измерительного инструмента имеет вид:

$$A_r = \begin{pmatrix} a_{11}^r & a_{12}^r & a_{13}^r \\ a_{21}^r & a_{22}^r & a_{23}^r \\ a_{31}^r & a_{32}^r & a_{33}^r \end{pmatrix},$$

где  $a_{ij}^r$  – количество переходов студентов из группы  $i$  в  $j$  после прохождения  $r$ -го инструмента.

Вероятности переходов студента из одной группы в другую определяются матрицей  $P^r$ , каждый элемент которой вычисляется по формуле:

$$p_{ij}^r = \frac{a_{ij}^r}{S_i^r}, \quad S_i^r = \sum_{j=1}^3 a_{ij}^r. \quad (1)$$

Таким образом, для курса, состоящего из  $n$  контрольно-измерительных инструментов, строится  $P^2, P^3, \dots, P^n$  матриц переходов, каждая из которых описывает дискретную цепь Маркова. Матрицу переходов по группам успешности, например, для модуля, включающего контрольно-измерительные инструменты от  $r_1$  до  $r_2$ , можно вычислить по формуле:

$$P = \prod_{i=r_1+1}^{r_2} P^i. \quad (2)$$

Имея матрицы перехода  $P^2, P^3, \dots, P^n$ , можно прогнозировать вероятность распределения студентов по группам успешности при

повторном прохождении курса (другими студентами или в следующем учебном году). Если после выполнения заданий  $r$ -ого контрольно-измерительного задания распределение студентов по группам соответствует вектору вероятности  $\vec{a}_r$ , можно вычислить вектор  $\vec{a}_k$  – вероятность распределения студентов после выполнения  $k$ -ого задания, используя соотношение:

$$\vec{a}_k = \left( \prod_{j=r+1}^k P^j \right) \cdot \vec{a}_r, \quad (3)$$

$$k = r + 1, \dots, n$$

### Результаты исследования

Для раннего прогнозирования успешности обучения классификация студентов осуществлялась после выполнения каждого контрольно-измерительного инструмента. В таблице 2 приведены результаты классификации в соотношении с группой успеваемости по одному из обучающих тестов самостоятельной работы. На рисунке 2а показана проекция классифицируемой выборки по двум признакам ( $f_2$  – результативность,  $f_3$  – периодичность), а на рисунке показан 1b результат обучения классификатора для этих же признаков.

Таблица 2

Результат классификации после выполнения обучающего теста

Table 2

### The classification prediction after completing a learning test

Группа	Всего студентов в группе успешности	Отл	Хор	Уд
1	229	137	73	19
2	180	56	92	32
3	127	19	49	61
Всего студентов в группе успеваемости		212	214	112

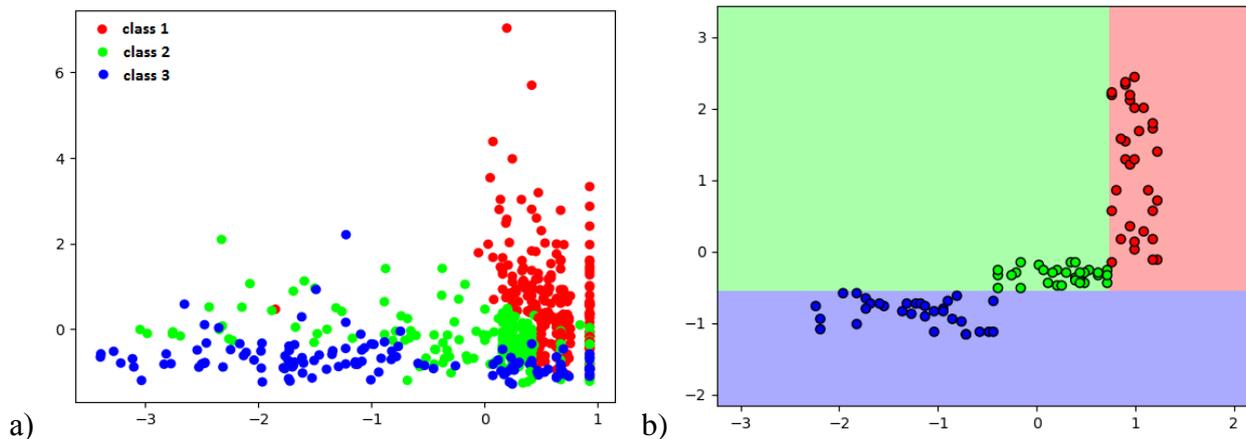


Рис. 2. а) Проекция классифицируемой выборки; б) Обучающая выборка

Fig. 2. a) The classification prediction; b) Training dataset

Из таблицы видно, что более 84 % студентов, попавших в 1 и 2 группы успешности, получили на экзамене 4 или 5. Причем группа с высокой успешностью состоит из 60 % отличников и 32 % хорошистов. Также необходимо отметить, что 88 % студентов, не сдавших экзамен с первого раза, попали в группу 3.

Мониторинг состава групп успешности со стороны преподавателя позволяет выявить потенциально «выпадающих» студентов за несколько месяцев до итоговой аттестации, когда ситуацию еще можно исправить. Все студенты, попавшие в 3-ю группу, входят в

группу «риска» (для нашего теста их 127 человек), о каждом из них формируется полный отчет об их деятельности на курсе и предоставляется преподавателю.

В таблице 3 приведен процент студентов, отнесенных классификатором к группе 3 с низким уровнем успешности в течение семестра определенное число раз, в соотношении с полученной оценкой по дисциплине. Из таблицы видно, что 95 % отчисленных студентов попадали в группу с низкой успешностью как минимум 6 раз, а вот только 16 % отличников оказывались в этой группе 1 или 2 раза.

Таблица 3

Результаты классификации для группы с низкой успешностью

Table 3

The classification prediction for a low successful group

Группа студентов по успеваемости	Процент студентов, попавших в группу указанное количество раз			
	1–2	3–5	6–10	Больше 10
Отчисленные студенты	100	100	95	89
Студенты, получившие «неуд»	100	92	75	71
Студенты, получившие «уд»	79	70	63	41
Студенты, получившие «хор»	45	38	21	10
Студенты, получившие «отл»	16	10	6	-

По окончании изучения модуля вновь проводится классификация студентов на группы успешности для каждой контрольной точки, чтобы учесть результаты тех попыток, которые студенты выполнили не в заданный преподавателем интервал времени. На основе этих данных, используя формулу 1, были построены матрицы переходов студентов по группам успешности ( $P^2, P^3, \dots, P^n$ ). На рисунке 3 показан граф вероятности переходов после выполнения обучающего теста в сравнении с предыдущим заданием – лабораторной

работой (рис. 2а). На рисунке видно, что студенты первой группы с вероятностью в 68 % в ней и остались, это хороший показатель. А вот тот факт, что 81 % студентов с низкой успешностью остались в своей группе, говорит о том, что с 3-ей группой студентов не было проведено необходимой работы со стороны преподавателя. Более 24 % студентов от общей численности понизили свою группу успешности (32 % студентов первой группы и 30 % второй) и только 10 % ее повысили (16 % студентов второй группы и 18 % третьей).

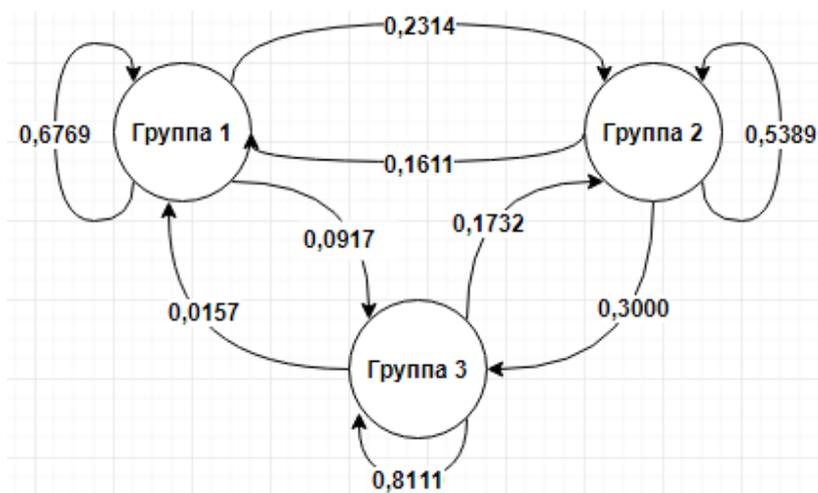


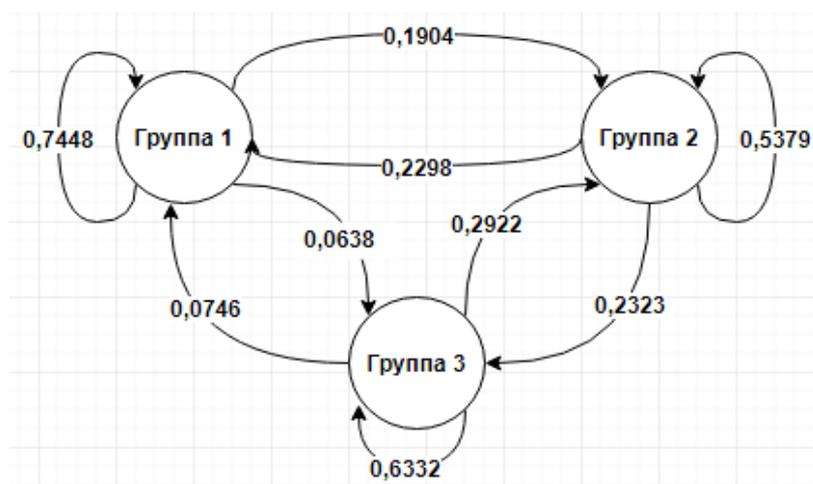
Рис. 3. Марковская цепь, для описания переходов между группами успешности по обучающему тесту

Fig. 3. Markov chain for transitions between learning success groups for training test

На основе этих данных можно предположить, что обучающий тест слишком труден для студентов, и преподавателю для следующих потоков студентов или в следующем учебном году следует либо добавить дополнительные разъяснения к заданиям, либо изменить его характеристики.

На рисунке 4 показан граф, построенный по матрице переходов, вычисленной по фор-

муле 2, для модуля. Модуль состоял из 10 контрольно-измерительных заданий. Анализируя граф, можно сделать вывод, что с вероятностями 75 %, 54 % и 63 % студенты сохраняют свою группу успешности. Для первой группы – это хороший показатель, а процент студентов, с низкой успешностью, требует дополнительных мер со стороны преподавателя.



**Рис. 4.** Марковская цепь, для описания переходов между группами успешности по модулю

**Fig. 4.** Markov chain for transitions between learning success groups for module

Наличие матриц перехода для каждого контрольно-измерительного инструмента за предыдущий учебный год позволяет выполнить прогнозирование успешности прохождения курса студентами в следующем учебном году. Так, имея вектор распределения студентов для  $i$ -го контрольно-измерительного инструмента текущего года, например,  $\vec{a}_i^T = (0,38 \ 0,41 \ 0,21)$ , и матрицу переходов  $P^{i+1}$  предыдущих лет, можно прогнозировать распределение студентов текущего года после его выполнения  $i + 1$  контрольно-измерительного инструмента. Вычисленный по формуле 3, вектор имеет вид:  $\vec{a}_{i+1}^T = (0,32 \ 0,35 \ 0,33)$ , то есть вероятно уменьшение численности первой и второй группы за счет увеличения третьей. Это не является желательным результатом, преподаватель может изменить сроки, сложность или другие параметры  $i + 1$ -го контрольно-измерительного инструмента.

### Заключение

Смешанное обучение представляет собой новую образовательную парадигму, технической составляющей которой является

наличие системы управления обучением (LMS), в которой могут быть организованы различные виды деятельности студентов: тестирование, самостоятельная работа, изучение теоретического материала и др. Все активности и достижения студента на online платформе оставляют «цифровой след» и фиксируются в LMS. Анализ таких данных (учебная аналитика) позволяет прогнозировать успешность обучения студента.

Проведенное исследование выявило следующие особенности прогнозирования:

- прогнозирование должно осуществляться на основе критериев, определяющих успешность обучения, таких как своевременность, результативность, непрерывность, самостоятельность; для всех факторов существуют метрики, построенные на основе данных учебной аналитики;

- классификацию студентов на группы успешности следует проводить для каждого контрольно-измерительного инструмента непосредственно после его выполнения студентами, чтобы своевременно выявить обучающихся «группы риска», нуждающихся в особом внимании со стороны преподавателя;



– необходимо накапливать информацию динамике переходов обучающихся между группами успешности, используя дискретные

цепи Маркова, что позволит прогнозировать успешность обучения других потоков студентов и в следующем учебном году.

## СПИСОК ЛИТЕРАТУРЫ

1. **Boelens R., De Wever B., Voet M.** Four key challenges to the design of blended learning: A systematic literature review // *Educational Research Review*. – 2017. – Vol. 22. – P. 1–18. DOI: <https://doi.org/10.1016/j.edurev.2017.06.001>
2. **Jones K. M.** Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy // *International Journal of Educational Technology in Higher Education*. – 2019. – Vol. 16 – Article number 24. DOI: <https://doi.org/10.1186/s41239-019-0155-0>
3. **Schumacher C., Ifenthaler D.** Features students really expect from learning analytics // *Computers in Human Behavior*. – 2018. – Vol. 78. – P. 397–407. DOI: <http://dx.doi.org/10.1016/j.chb.2017.06.030>
4. **Nistor N., Hernández-García Á.** What types of data are used in learning analytics? An overview of six cases // *Computers in Human Behavior*. – 2018. – Vol. 89. – P. 335–338. DOI: <https://doi.org/10.1016/j.chb.2018.07.038>
5. **Быстрова Т. Ю., Ларионова В. А., Сеницын Е. В., Толмачев А. В.** Учебная аналитика MOOC как инструмент прогнозирования успешности обучающихся // *Вопросы образования*. – 2018. – № 4. – С. 139–166. DOI: <https://doi.org/10.17323/1814-9545-2018-4-139-166> URL: <https://elibrary.ru/item.asp?id=36566170>
6. **Патаракин Е. Д.** Совместная сетевая деятельность и поддерживающая ее учебная аналитика // *Высшее образование в России*. – 2015. – № 5. – С. 145–154. URL: <https://elibrary.ru/item.asp?id=23726842>
7. **Lim L. A., Gentili S., Pardo A., Kovanović V., Whitelock-Wainwright A., Gašević D., Dawson S.** What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course // *Learning and Instruction*. – 2019. DOI: <https://doi.org/10.1016/j.learninstruc.2019.04.003>
8. **Бирина О. В.** Понятие успешности обучения в современных педагогических и психологических теориях // *Фундаментальные исследования*. – 2014. – № 8-2. – С. 438–443. URL: <https://elibrary.ru/item.asp?id=21736773>
9. **Gardner J., Brooks C.** Student success prediction in MOOCs // *User Modeling and User Adapted Interaction*. – 2018. – Vol. 28, Issue 2. – P. 127–203. DOI: <http://dx.doi.org/10.1007/s11257-018-9203-z>
10. **Raga R., Raga J.** Early prediction of student performance in blended learning courses using deep neural networks International // *Symposium on Educational Technology*. – 2019. – P. 39–43. DOI: <https://doi.org/10.1109/ISET.2019.00018>
11. **Xu X., Wang J., Peng H., Wu R.** Prediction of academic performance associated with internet usage behaviors using machine learning algorithms // *Computers in Human Behavior*. – 2019. – Vol. 98. – P. 166–173. DOI: <https://doi.org/10.1016/j.chb.2019.04.015>
12. **Moreno-Marcos P. M., Alario-Hoyos C., Munoz-Merino P. J., Kloos C. D.** Prediction in MOOCs: A Review and Future Research Directions // *Transactions on Learning Technologies*. – 2019. – Vol. 12, Issue 3. – P. 384–401. DOI: <https://doi.org/10.1109/TLT.2018.2856808>



13. **Ellis R. A., Pardo A., Han F.** Quality in blended learning environments—Significant differences in how students approach learning collaborations // *Computers and Education*. – 2016. – Vol. 102. – P. 90–102. DOI: <https://doi.org/10.1016/j.compedu.2016.07.006>
14. **Conijn R., Van den Beemt A., Cuijpers P.** Predicting student performance in a blended MOOC // *Journal of Computer Assisted Learning*. – 2018 – Vol. 34, Issue 5. – P. 615–628. DOI: <https://doi.org/10.1111/jcal.12270>
15. **Kim D., Yoon M., Jo I. H., Branch R. M.** Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women’s university in South Korea // *Computers and Education*. – 2018. – Vol. 127. – P. 233–255. DOI: <https://doi.org/10.1016/j.compedu.2018.08.023>
16. **Conijn R., Snijders C., Kleingeld A., Matzat U.** Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle // *Transactions on Learning Technologies*. – 2017. – Vol. 10, № 1. – P. 17–29. DOI: <https://doi.org/10.1109/TLT.2016.2616312>
17. **Fei M., Yeung D.** Temporal Models for Predicting Student Dropout in Massive Open Online Courses // *International Conference on Data Mining Workshop*. – 2015. – P. 256–263. DOI: <https://doi.org/10.1109/ICDMW.2015.174>
18. **Macina J., Srba I., Williams J. J., Bielikova M.** Educational question routing in online student communities // *ACM Conference on Recommender Systems*. – 2017. – P. 47–55. DOI: <https://doi.org/10.1145/3109859.3109886>
19. **Topîrceanua A., Grosseckb G.** Decision tree learning used for the classification of student archetypes in online courses // *Procedia Computer Science*. – 2017. – Vol. 112. – P. 51–60. DOI: <https://doi.org/10.1016/j.procs.2017.08.021>
20. **Xing W., Chen X., Stein J., Marcinkowski M.** Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization // *Computers in Human Behavior*. – 2016. – Vol. 58. – P. 119–129. DOI: <https://doi.org/10.1016/j.chb.2015.12.007>
21. **Laveti R. N., Kuppili S., Ch J., Pal S. N., Babu N. S. C.** Implementation of learning analytics framework for MOOCs using state-of-the-art in-memory computing // *National Conference on E-Learning and E-Learning Technologies*. – 2017. – P. 1–6. DOI: <https://doi.org/10.1109/ELELTECH.2017.8074997>
22. **Gamulin J., Gamulin O., Kermek D.** Using fourier coefficients in time series analysis for student performance prediction in blended learning environments // *Expert Systems*. – 2016. – Vol. 33, № 2. – P. 189–200. DOI: <https://doi.org/10.1111/exsy.12142>
23. **Nam S., Frishkoff G., Collins-Thompson K.** Predicting Students Disengaged Behaviors in an Online Meaning-Generation Task // *IEEE Transactions on Learning Technologies*. – 2018. – Vol. 11, № 3. – P. 362–375. DOI: <https://doi.org/10.1109/TLT.2017.2720738>
24. **Arabshahi F., Huang F., Anandkumar A., Butts C. T., Fitzhugh S. M.** Are You Going to the Party: Depends, Who Else is Coming? [Learning Hidden Group Dynamics via Conditional Latent Tree Models] // *International Conference on Data Mining*. – 2015. – P. 697–702. DOI: <https://doi.org/10.1109/ICDM.2015.146>
25. **Youssef M., Mohammed S., Hamada E. K., Wafaa B. F.** A predictive approach based on efficient feature selection and learning algorithms’ competition: Case of learners’ dropout in MOOCs // *Education and Information Technologies*. – 2019. – Vol. 24, Issue 6. – P. 3591–3618. DOI: <https://doi.org/10.1007/s10639-019-09934-y>
26. **Шестакова Л. Г., Рихтер Т. В.** Показатели оценки и самооценки готовности студентов к самоорганизации // *Science for Education Today*. – 2019. – Т. 9, № 3. – С. 138–150. DOI: <http://dx.doi.org/10.15293/2658-6762.1903.08> URL: <https://elibrary.ru/item.asp?id=38538212>



27. **Klein C., Lester J., Rangwala H., Johri A.** Learning Analytics Tools in Higher Education: Adoption at the Intersection of Institutional Commitment and Individual Action // Review of Higher Education. – 2019. – Vol. 42, № 2. – P. 565–593. DOI: <https://doi.org/10.1353/rhe.2019.0007>



DOI: [10.15293/2658-6762.1906.05](https://doi.org/10.15293/2658-6762.1906.05)

Galina Pavlovna Ozerova,

Candidate of Technical Science, Associate Professor,  
Department of Mechanics and Mathematical Modeling, Engineering school,  
Far Eastern Federal University, Vladivostok, Russian Federation.

Corresponding author

ORCID ID: <https://orcid.org/0000-0002-1630-3888>

E-mail: [ozerovalp@dvfu.ru](mailto:ozerovalp@dvfu.ru)

Galina Fedorovna Pavlenko,

Candidate of Technical Science, Associate Professor,  
Department of Business Informatics and Economic and  
Mathematical Methods, School of Economics and Management,  
Far Eastern Federal University, Vladivostok, Russian Federation.

ORCID ID: <https://orcid.org/0000-0002-1637-642X>

E-mail: [pavlenko.gf@dvfu.ru](mailto:pavlenko.gf@dvfu.ru)

## Prediction of student performance in blended learning utilizing learning analytics data

### Abstract

**Introduction.** This paper is devoted to predicting performance of students involved in blended learning. The objective of the research is to identify characteristics of predicting student performance in blended learning using learning analytics data.

**Materials and Methods.** Primary methods used in the research are the following: theoretical analysis and generalization of previous studies, theoretical and practical methods of educational research, statistical processing of empirical data, machine learning and random events modelling.

**Results.** The research has found that predication has to be based on the criteria which determine learning success. Metrics for the criteria can be obtained through learning analytics data. Students should be split into groups according to their academic performance every time they complete their assignments in order to identify low performers who require support from academic staff. In order to predict future performance more efficiently, we need to accumulate dynamics of how students get re-classified into groups using discrete Markov Chains.

**Conclusions.** Prediction of student performance based on learning analytics data allows to identify students who fall into high risk group, predict how students can be distributed among performance groups, and adopt teaching materials to student needs.

### Keywords

Learning success; Blended learning; Learning management system; Learning analytics; Prediction; Classification; Discrete Markov Chains.

## REFERENCES

1. Boelens R., De Wever B., Voet M. Four key challenges to the design of blended learning: A systematic literature review. *Educational Research Review*, 2017, vol. 22, pp. 1–18. DOI: <https://doi.org/10.1016/j.edurev.2017.06.001>
2. Jones K. M. Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of*



- Educational Technology in Higher Education*, 2019, vol. 16, article number 24. DOI: <https://doi.org/10.1186/s41239-019-0155-0>
3. Schumacher C., Ifenthaler D. Features students really expect from learning analytics. *Computers in Human Behavior*, 2018, vol. 78, pp. 397–407. DOI: <http://dx.doi.org/10.1016/j.chb.2017.06.030>
  4. Nistor N., Hernández-García Á. What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior*, 2018, vol. 89, pp. 335–338. DOI: <https://doi.org/10.1016/j.chb.2018.07.038>
  5. Bystrova T. Yu., Larionova V. A., Sinitsyn E. V., Tolmachev A. V. Learning analytics in massive open online courses as a tool for predicting learner performance. *Voprosy obrazovaniya / Educational Studies Moscow*, 2018, no. 4, pp. 139–166. DOI: <https://doi.org/10.17323/1814-9545-2018-4-139-166> URL: <https://elibrary.ru/item.asp?id=36566170>
  6. Patarakin E. D. Learning analytic to facilitate group network collaboration. *Vysshee Obrazovanie v Rossii (Higher Education in Russia)*, 2015, no. 5, pp. 145–154. (in Russian) URL: <https://elibrary.ru/item.asp?id=23726842>
  7. Lim L. A., Gentili S., Pardo A., Kovanović V., Whitelock-Wainwright A., Gašević D., Dawson S. What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, 2019. DOI: <https://doi.org/10.1016/j.learninstruc.2019.04.003>
  8. Birina O. V. The concept of successful learning in modern pedagogical and psychological theories. *Fundamental Research*, 2014, no. 8-2, pp. 438–443. (in Russian) URL: <https://elibrary.ru/item.asp?id=21736773>
  9. Gardner J., Brooks C. Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 2018, vol. 28, issue 2, pp. 127–203. DOI: <http://dx.doi.org/10.1007/s11257-018-9203-z>
  10. Raga R., Raga J. Early prediction of student performance in blended learning courses using deep neural networks International. *Symposium on Educational Technology*, 2019, pp. 39–43. DOI: <https://doi.org/10.1109/ISET.2019.00018>
  11. Xu X., Wang J., Peng H., Wu R. Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 2019, vol. 98, pp. 166–173. DOI: <https://doi.org/10.1016/j.chb.2019.04.015>
  12. Moreno-Marcos P. M., Alario-Hoyos C., Muñoz-Merino P. J., Kloos C. D. Prediction in MOOCs: A review and future research directions. *Transactions on Learning Technologies*, 2019, vol. 12, issue 3, pp. 384–401. DOI: <https://doi.org/10.1109/TLT.2018.2856808>
  13. Ellis R. A., Pardo A., Han F. Quality in blended learning environments—Significant differences in how students approach learning collaborations. *Computers and Education*, 2016, vol. 102, pp. 90–102. DOI: <https://doi.org/10.1016/j.compedu.2016.07.006>
  14. Conijn R., Van den Beemt A., Cuijpers P. Predicting student performance in a blended MOOC. *Journal of Computer Assisted Learning*, 2018, vol. 34, issue 5, pp. 615–628. DOI: <https://doi.org/10.1111/jcal.12270>
  15. Kim D., Yoon M., Jo I. H., Branch R. M. Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women’s university in South Korea. *Computers and Education*, 2018, vol. 127, pp. 233–255. DOI: <https://doi.org/10.1016/j.compedu.2018.08.023>
  16. Conijn R., Snijders C., Kleingeld A., Matzat U. Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle. *Transactions on Learning Technologies*, 2017, vol. 10, issue 1, pp. 17–29. DOI: <https://doi.org/10.1109/TLT.2016.2616312>



17. Fei M., Yeung D. Temporal models for predicting student dropout in massive open online courses. *International Conference on Data Mining Workshop*, 2015, pp. 256–263. DOI: <https://doi.org/10.1109/ICDMW.2015.174>
18. Macina J., Srba I., Williams J. J., Bielikova M. Educational question routing in online student communities. *ACM Conference on Recommender Systems*, 2017, pp. 47–55. DOI: <https://doi.org/10.1145/3109859.3109886>
19. Topîrceanua A., Grossecckb G. Decision tree learning used for the classification of student archetypes in online courses. *Procedia Computer Science*, 2017, vol. 112, pp. 51–60. DOI: <https://doi.org/10.1016/j.procs.2017.08.021>
20. Xing W., Chen X., Stein J., Marcinkowski M. Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 2016, vol. 58, pp. 119–129. DOI: <https://doi.org/10.1016/j.chb.2015.12.007>
21. Laveti R. N., Kuppili S., Ch J., Pal S. N., Babu N. S. C. Implementation of learning analytics framework for MOOCs using state-of-the-art in-memory computing. *National Conference on E-Learning and E-Learning Technologies*, 2017, pp. 1–6. DOI: <https://doi.org/10.1109/ELELTECH.2017.8074997>
22. Gamulin J., Gamulin O., Kermek D. Using fourier coefficients in time series analysis for student performance prediction in blended learning environments. *Expert Systems*, 2016, vol. 32, issue 2, pp. 189–200. DOI: <https://doi.org/10.1111/exsy.12142>
23. Nam S., Frishkoff G., Collins-Thompson K. Predicting Students disengaged behaviors in an online meaning-generation task. *IEEE Transactions on Learning Technologies*, 2018, vol. 11, issue 3, pp. 362–375. DOI: <https://doi.org/10.1109/TLT.2017.2720738>
24. Arabshahi F., Huang F., Anandkumar A., Butts C. T., Fitzhugh S. M. Are you going to the party: Depends, who else is coming? [Learning hidden group dynamics via conditional latent tree models]. *International Conference on Data Mining*, 2015, pp. 697–702. DOI: <https://doi.org/10.1109/ICDM.2015.146>
25. Youssef M., Mohammed S., Hamada E. K., Wafaa B. F. A predictive approach based on efficient feature selection and learning algorithms' competition: Case of learners' dropout in MOOCs. *Education and Information Technologies*, 2019, vol. 24, issue 6, pp. 3591–3618. DOI: <https://doi.org/10.1007/s10639-019-09934-y>
26. Shestakova L. G., Richter T. V. Student's readiness for self-organization: Indicators of evaluation and self-evaluation. *Science for Education Today*, 2019, vol. 9, issue 3, pp. 138–150. DOI: <http://dx.doi.org/10.15293/2658-6762.1903.08> URL: <https://elibrary.ru/item.asp?id=38538212>
27. Klein C., Lester J., Rangwala H., Johri A. Learning analytics tools in higher education: Adoption at the intersection of institutional commitment and individual action. *Review of Higher Education*, 2019, vol. 42, issue 2, pp. 565–593. DOI: <https://doi.org/10.1353/rhe.2019.0007>

Submitted: 09 October 2019

Accepted: 08 November 2019

Published: 31 December 2019



This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. (CC BY 4.0).