



Анализ и кластеризация основных факторов, влияющих на успеваемость учебных групп вуза

М.В. Гранков, В.М. Аль-Габри, М.Ю. Горлова

Донской государственный технический университет, Ростов-на-Дону

Аннотация: В статье рассматриваются подходы к задаче повышения успеваемости студентов вузов. Под успеваемостью понимается процент положительных оценок, полученных студентами по результатам сессий, без учета пересдач. Анализ факторов, влияющих на успеваемость, рассматривается на примере информации из базы данных оценок портала УМУ Донского государственного технического университета. Выделены основные факторы (объекты), влияющие на успеваемость: учебные группы, преподаватели, предметы, расписание экзаменов. Применение методов кластеризации позволило использовать в настоящей работе дисперсионный анализ средних. С помощью двухфакторного дисперсионного анализа установлено существенное влияние классов объектов «Преподаватель» и «Предмет» на процент положительных оценок, полученных студентами по результатам сессии. Приведен пример классов групп, преподавателей и предметов, на результаты экзаменов которых оказали существенное влияние расписания проведения экзаменов.

Ключевые слова: высшее образование, успеваемость, дисперсионный анализ, факторы, кластеризация.

Введение: Современное состояние российской образовательной системы отличает длительный период реформирования, обусловленный попытками федерального центра преодолеть традиционную инерционность этой системы, препятствующую России занять достойное место в рейтинге стран, разрабатывающих и внедряющих современные наукоемкие технологии. Известные процессы, предшествующие и последующие за распадом СССР до сих пор оказывают отрицательное влияние, приводящее к снижению качества образования в Российской Федерации. Одним из результатов таких деструктивных процессов является значительное снижение уровня успеваемости студенческих групп, рассчитанного на конец сессии. В данном случае речь идет о проценте положительных оценок, полученных студентами некоторого множества учебных групп по результатам экзаменационной сессии. Снижение этого процента повышает общее количество задолженностей по вузу, что приводит к увеличению «неофициальной»

компоненты нагрузки преподавателей, связанной с ликвидацией задолженностей студентов по уже изученным дисциплинам. Это создает условия совпадения интересов студентов, преподавателей и администрации вузов, что приводит, в лучшем случае, к снижению уровня освоения студентами компетенций, предусмотренных образовательными программами. Проблеме повышения и анализу успеваемости студенческих групп посвящены научно-исследовательские и прикладные работы [2-4,13,14], а кластерному анализу показателей образовательной деятельности студентов вузов статьи [11,12]. В связи с важностью формальных исследований, позволяющих не только повысить качество образовательных процессов в вузе, но и снизить рутинную нагрузку на преподавателей, актуальными являются работы, связанные с анализом факторов, влияющих на уровень средней успеваемости студенческих групп по результатам сессий. В настоящем исследовании делается попытка анализа и обоснования основных факторов, влияющих на уровень успеваемости студенческих групп по результатам экзаменационных сессий.

Цель: Разработать метод, позволяющий проанализировать и формально обосновать степень влияния основных факторов, участвующих в формировании результатов экзаменов, на общую успеваемость студентов вуза.

Исходные данные и основные объекты исследования: В данном исследовании использовалась база данных электронных деканатов Донского Государственного Технического Университета. Эта база данных поддерживается с 1999 года и содержит достаточно полную информацию об итогах зачетных и экзаменационных сессий студентов факультетов ДГТУ. На первом этапе внедрения программного обеспечения электронных деканатов информация о результатах испытаний вводилась сотрудниками деканатов, без указания индивидуальной даты получения положительной оценки



каждым студентом. Возможность регистрации даты получения каждым студентом положительной оценки на основном экзамене, а, следовательно, и расчет процента успеваемости группы на конец сессии, появилась с внедрением в 2007 году программного обеспечения, рассчитанного на индивидуальный ввод в электронные ведомости каждым преподавателем информации о результатах проведенного им экзамена.

Предварительная обработка исходных данных: Первый этап обработки данных заключался в выделении индивидуальных экзаменационных оценок студентов, поступивших в ДГТУ с сентября 2007/2008 учебного года. Для выделенной таким образом информации, был построено отношение результатов экзаменов учебных групп с данными об успеваемости по каждой дисциплина на момент окончания соответствующих сессий. В это отношение включалась информация только о результатах экзаменов студентов головного университета, обучавшихся по образовательным программам специалистов и бакалавров, дневной формы обучения, сдававших экзамены в периоды соответствующих сессий. Обработка проводилась на специально разработанном программном стенде, при этом, исходное отношение содержало 2697797 записей о результатах индивидуальных испытаний студентов, а результирующее отношение успеваемости учебных групп содержал 20511 записей. Логический блок стенда позволял обнаруживать противоречия в информации исходном отношении индивидуальных результатов испытаний и принимать решений о корректировке исходной информации, либо протоколировать обнаруженные проблемы. Построенное результирующее отношение «Результаты экзаменов учебных групп» находится в первой нормальной форме и содержит избыточную информацию. Однако, эта избыточность позволяет легко реализовать исследования успеваемости учебных групп в различных разрезах: по годам, курсам, семестрам, сессиям, специальностям, предметам, преподавателям и

т.д. К полученным выборкам могут быть применены методы статистической обработки и анализа.

Модель успеваемости учебных групп по результатам сессий:

На результаты экзаменов оказывают влияние большое число объектов, процессов, факторов и условий различного происхождения и природы. В реализации экзаменов участвуют группы людей, что существенно усложняет применение стандартного аппарата параметрических статистик для анализа результатов их действий. В [2-3] отмечается известный в психологии факт, что реальные законы, управляющие поведением людей в группах, сложнее, чем описываемые нормальным законом распределения. Однако, рассматривая учебную группу как некий интегратор этих процессов, можно предположить, что такие группы, как пример социальных групп, могут быть классифицированы и иметь, в зависимости от внешних воздействий, достаточно предсказуемое поведение. Опираясь на это предположение, выделим основные типы объектов, влияющих на формирование значения общего числа задолженностей в вузе по результатам сессии:

1. учебная группа;
2. экзаменатор (преподаватель);
3. предмет (дисциплина);
4. расписание экзаменов (сессии).

В настоящем исследовании под расписанием экзаменов будем понимать некоторую последовательность дисциплин, в которой группа сдавала экзамены преподавателям по соответствующим дисциплинам. Формально расписание экзаменов это отношение, построенное на прямом произведении множеств объектов: учебная группа, экзаменатор, дисциплина. Представим учебную группу некоторым объектом, преобразующим внешние воздействия экзаменатора, дисциплины и расписания сессии в значение успеваемости. Введем следующие обозначения:



$$N = \{1, 2, \dots, i, \dots\} \quad (1)$$

— множество натуральных чисел.

$$N_0 = \{0\} \cup N \quad (2)$$

— расширенное множество натуральных чисел.

$$I = \{0, 1\} \quad (3)$$

— единичный отрезок действительных чисел.

$$E_n = \left\{ r_i \mid (i = 1, 2, \dots, n) \wedge \left(r_i = \frac{n1_i}{n2_i} \right) \wedge (n1_i \in N_0) \wedge (n2_i \in N) \wedge (n1_i \leq n2_i) \wedge (n \in N) \right\} \quad (4)$$

— множество неотрицательных рациональных чисел, представленных дробями, значение которых не превышает 1.

$$G = \{1, 2, \dots, n_g\}, n_g \in N \quad (5)$$

— множество учебных групп, где n_g – число учебных групп.

$$T = \{1, 2, \dots, n_t\}, n_t \in N \quad (6)$$

— множество экзаменаторов (преподавателей), где n_t – число экзаменаторов.

$$S = \{1, 2, \dots, n_s\}, n_s \in N \quad (7)$$

— множество дисциплин (предметов), где n_s – число дисциплин.

$$TT \subset G \times N \times N \times N \times T \times S \times N_0 \times N \times U_n \quad (8)$$

где отношение TT содержит результаты экзаменов групп вуза на конец соответствующих сессий. Кортеж $(g, y, i, j, t, s, n1, n2, u) \in TT$, если группа $g \in G$, год поступления в вуз которой равен $y \in N$, сдавала в сессию под номером $i \in N$ экзамен под номером $j \in N$ преподавателю $t \in T$ по дисциплине $s \in S$. Значение $n2$ соответствует числу студентов, которые должны были сдавать данный экзамен, а $n1$ соответствует числу студентов из $n2$, получивших положительные оценки. Справочное значение $u (u \in E_n)$, является успеваемостью группы по результатам данного экзамена.

Пусть TT_1 – некоторое подмножество отношения $TT (TT_1 \subset TT)$, которое соответствует некоторой выборке из последнего отношения. В зависимости от запросов в подмножество TT_1 могут включаться результаты экзаменов одной или нескольких групп. Рассмотрим проекцию множества TT_1 по атрибутам 7 и 8 (число положительных оценок и общее число студентов, соответственно).

$$\pi_{7,8}(TT_1) = \{(n1_1, n2_1), (n1_2, n2_2), \dots, (n1_k, n2_k)\}, \text{ где} \quad (9)$$

$\pi_{7,8}(\ast)$ – реляционная операция проектирования отношения по атрибутам под номерами семь и восемь;

$k = |TT_1|$ – общее число результатов экзаменов, включенных в выборку.

В этом случае среднюю успеваемость на экзаменах групп, включенных в выборку, можно найти по формуле:

$$u_1 = \left(\frac{\sum_{i=1}^n n1_i}{\sum_{i=1}^n n2_i} \right), \quad (10)$$

очевидно, что $u_1 \in E_n$.

Задача и методы кластеризации объектов:

Одним из стандартных аппаратов решения проблем, подобных сформулированной в цели настоящего исследования, является дисперсионный анализ [6-10]. Основная и популярная у исследователей часть этого аппарата базируется на гипотезах о нормальности распределений, однородности математических ожиданий и дисперсий, что очень проблематично в исследуемой предметной области. Как известно, множество полученных отдельным студентом оценок подчиняется непараметрической статистике. В работах [2-3] рассмотрены проблемы, которые возникают при переходе от непараметрической статистики успеваемости каждого студента к параметрической статистике успеваемости в больших выборках студентов. В

этих исследованиях под успеваемостью понималось среднее значение оценок (баллов), которые студент набрал за некоторый интервал времени обучения. Было обнаружено, что на границах допустимых значений оценок возникают функциональные зависимости между средней успеваемостью в локальном интервале и средней дисперсией, рассчитанной в том же интервале. При этом отмечалось влияние на данную зависимость процедуры пересдач неудовлетворительных оценок. В результате, средние успеваемости, полученные для больших выборок студентов, могут не подчиняться нормальному закону распределения случайной величины, даже если для данного распределения выполнен критерий Хи-квадрат.

В настоящем исследовании, в отличие от работ[2-4]:

1. понятие успеваемости рассматривается только как доля студентов, получивших положительные оценки на конец сессии, к общему числу студентов, которые должны были сдавать экзамены;
2. значение показателя успеваемости регистрируется на конец сессии, без учета пересдач;
3. основным объектом исследования является не студент, а учебная группа. При этом, результатом измерения является успеваемость учебной группы, показанная по дисциплине на экзамене, проведенным в сессию (в соответствии с расписанием) некоторым преподавателем.

В общем случае на результат измерения успеваемости оказывают влияние: учебная группа, дисциплина, преподаватель, а, возможно, и расписание экзаменов [1]. Такое количество факторов значительно усложняет задачу анализа успеваемости. Воспользуемся аппаратом дисперсионного анализа средних, корректность которого доказана в [9], в том числе для случая неравенства дисперсий и неравных чисел измерений в ячейках. Нами предложено получать значения факторов, которые необходимы для реализации дисперсионного анализа средних, с помощью методов

кластеризации. Для этого, из выделенных нами ранее основных типов объектов образуем три типа качественных факторов (группа, преподаватель, предмет), влияющих на число задолженностей в вузе по результатам сессии. Соответствующие факторам объекты могут распадаться на подмножества, влияющие близким образом на результаты экзаменов. Для оценки влияния основных факторов на среднюю успеваемость с помощью дисперсионного анализа формализуем механизм деления факторов (объектов) на классы, используя методы нечеткой кластеризации. Для этого, каждому объекту из типов: группа, преподаватель, предмет, поставим в соответствие количественное значение фактора. В качестве такого значения используем значение средней успеваемости. В нашем случае измерение значения средней успеваемости реализуется проведением экзаменов в соответствии с расписанием сессии. Так как, число студентов в учебных группах могут различаться, то в общем случае мы имеем дело с неравноточными измерениями. Тогда, под средней успеваемостью следует понимать средневзвешенную успеваемость [6-7]. Формула (10) является рациональным вариантом расчета средневзвешенной успеваемости, когда под весом понимается количество студентов, которые должны были сдавать экзамен. При этом, для одной группы формула (10) позволяет вычислить среднюю успеваемость, а для множеств групп, преподавателей или предметов – средневзвешенную успеваемость. В дальнейших рассуждениях под средней успеваемостью, вычисленную по формуле (10), будем понимать средневзвешенную успеваемость, как более общий ее вид. Очевидно, что полученные средние значения для соответствующих факторов (объектов) могут проявляться с той, или иной степенью стабильности. Для оценки стабильности средней успеваемости будем использовать средневзвешенную выборочную дисперсию [6]. Следовательно, каждый рассматриваемый объект можно описать двумерным вектором x :

$$x = (u, D_u), \quad (11)$$

где: u – выборочная средняя успеваемость;

D_u – выборочная средняя дисперсия успеваемости.

Как известно, декомпозиция объектов на классы должна обладать двум свойствами [4,5,11,12]:

- *гомогенность объектов в классах*, т.е. объекты в пределах одного и того же класса должны быть как можно более подобными друг другу;
- *гетерогенность между классами*, т.е. объекты, принадлежащие к разным группам, должны быть как можно сильнее отличаться друг от друга.

Каждый, рассматриваемый нами объект (фактор) представим вектором (17) в двумерном пространстве свойств объектов. В качестве меры подобия объектов используются различные метрики. Наиболее популярным способом измерения подобия объектов считается метрика Евклида. Одним из распространенных алгоритмов нечеткой кластеризации является алгоритм FCM (*Fuzzy C-Means*), который позволяет приписать одни и те же объекты к разным классам с соответствующими степенями принадлежности [5].

Результаты кластеризации:

За срок жизни учебная группа сдает порядка 30-40 экзаменов по предметам различной сложности и у различных преподавателей. Разделим учебные группы на пять классов по параметру «средняя успеваемость». Первоначальные эксперименты с одномерной декомпозицией учебных групп на классы по первому параметру вектора x (11) показали, что добиться однородности по оценкам успеваемости и ее дисперсии для всех групп, входящих в один класс, не представляется возможным. Удовлетворительный результат был получен после применения двух этапной кластеризации. На первом этапе – одномерная кластеризация на пять классов по средней успеваемости, а на втором этапе – двумерная кластеризация на два – пять

подклассов по параметрам вектора x (11), проведенная для множеств групп каждого класса, полученных на первом этапе. Объекты типов «Преподаватели» и «Предметы» были разбиты на пять классов каждый. Проверка однородности по оценке дисперсии в полученных классах не производилась. Классы всех типов объектов упорядочены по убыванию успеваемости.

Анализ влияния факторов на успеваемость учебных групп:

Приведем модифицированный алгоритм двухфакторного дисперсионного анализа средних, корректность которого доказана в [9]:

1. разобьем каждое из множеств преподавателей и предметов на пять классов;
 2. заполним таблицу средних следующим образом: число u_{ij} , стоящее на пересечении i -й строки и j -го столбца таблицы – это средняя взвешенная успеваемость, показанная в учебных группах по результатам экзаменов по предмету класса i , и проводимых преподавателями класса j . Удалим из таблицы пятую строку и пятый столбец, т.к. в них содержались нулевые ячейки [9], получая таблицу 1.
 3. Обработаем данные из таблицы 1, получая общую дисперсию $D_{общ}$ и групповые дисперсии D_1 и D_2 по факторам «Преподаватели» «Предметы» [10].
 4. Высказываем две нулевых гипотезы H_{01} и H_{02} о несущественности отличий групповых математических ожиданий успеваемости соответственно по первому фактору («Преподаватели») и по второму («Предметы»). При проверке гипотез используются критерии (12) для гипотезы H_{01} и (13) гипотезы H_{02} .
 5. Высказываем нулевую гипотезу H_{03} о несущественности различий в степени влияния значений первого и второго фактора на результаты измерений. При проверке гипотезы используется критерий (14).
-

Таблица № 1.

Таблица средних значений успеваемости в зависимости от уровней двух факторов: «Преподаватели» и «Предметы»

Класс предмета	Класс преподавателя				
	1	2	3	4	$\overline{u1}_i^*$
1	0.889756	0.8183	0.7883	0.62727	0.7809065
2	0.834071	0.679	0.5564	0.44228	0.627938
3	0.791342	0.616	0.4742	0.34155	0.555773
4	0.719231	0.3986	0.3504	0.23529	0.42588
$\overline{u2}_j^*$	0.8086	0.627975	0.542325	0.4115975	<u>0.59762</u>

$$f_1 = \frac{D_1^*}{D_{осм}} \quad (12)$$

— значение критерия для нулевой гипотезы H_{01}

$$f_2 = \frac{D_2^*}{D_{осм}} \quad (13)$$

— значение критерия для нулевой гипотезы H_{02}

$$f_3 = \frac{\max(D_1^*, D_2^*)}{\min(D_1^*, D_2^*)} \quad (14)$$

— значение критерия для нулевой гипотезы H_{03}

Результаты выполненных расчетов, сведены в таблицу 2

Таблица № 2

Результаты двухфакторного дисперсионного анализа успеваемости от классов объектов «Преподаватели» и «Предметы»

Название	D_1^*	D_2^*	$D_{осм}$	f_1	$F_1(\infty, 3)$	f_2	$F_2(\infty, 3)$	f_3	$F_3(3, 3)$
Значение	0.332	0.263	0.03	11.1	2.6	8.87	2.6	1.26	9.28

В таблице 2 значения $F_1(\infty,3)$, $F_2(\infty,3)$ и $F_3(3,3)$ – критические значения $f_{0.95}$ для распределения Фишера. Значение первого аргумента $F_1(\infty,3)$ и $F_2(\infty,3)$, согласно [9], соответствует значению степени свободы, связанное с числом измерений успеваемости каждой группы на каждом экзамене. Число таких измерений в проведенном исследовании значительно превышает максимальное число степеней свободы, предусмотренной в таблицах распределения Фишера. В связи с этим, значение первого параметра для таблиц распределения Фишера было принято равным бесконечности.

Из анализа Таблицы 2 следует:

1. Значение критерия f_1 значительно превышает соответствующее ему критическое значение $F_1(\infty,3)$ из таблиц распределения Фишера. Следовательно, нулевая гипотеза H_{01} не может быть принята и успеваемость групп по результатам экзаменов существенно зависит от класса преподавателей. Т.е. успеваемость группы по результатам экзамена у некоторого преподавателя существенно зависит от средней успеваемости, показанной различными группами, на экзаменах по различным предметам у этого преподавателя.
 2. Значение критерия f_2 значительно превышает соответствующее ему критическое значение $F_2(\infty,3)$ из таблиц распределения Фишера. Следовательно, нулевая гипотеза H_{02} не может быть принята и успеваемость групп по результатам экзаменов существенно зависит от класса предметов. Т.е. успеваемость группы по результатам экзамена по некоторому предмету существенно зависит от средней успеваемости, показанной различными группами, на экзаменах по этому предмету у разных преподавателей.
 3. Значение критерия f_3 значительно меньше соответствующего ему
-

критическое значение $F_3(3,3)$ из таблиц распределения Фишера. Следовательно, нулевая гипотеза H_{03} может быть принята и успеваемость по результатам экзаменов, проводимых преподавателями различных классов, не зависит существенно от классов предметов этих экзаменов

Для проверки гипотезы о влиянии расписания на общее число положительных оценок, полученных группой по результатам экзаменов сессии [1], был реализован однофакторный дисперсионный анализ. Под количественным значением фактора расписания будем понимать номер экзамена. С помощью программного стенда выберем измерения успеваемости для групп, сессии которых содержали не более четырех экзаменов. В результате получаем задачу с четырьмя уровнями фактора «Расписание». Применим в этой задаче дисперсионный анализ средних значений с различным числом измерений в ячейках [7-10]. В настоящей работе исследовалось влияния расписания экзаменов на успеваемость групп второго класса, которые сдавали предметы второго класса преподавателям, отнесенных при кластеризации ко второму классу. Опуская промежуточные расчеты, приведем в Таблице 3 окончательные результаты.

Таблица № 3.

Результаты анализа влияния расписания на успеваемость

Название	$D_{мг}^*$	$D_{вг}^*$	f	$F(N-1, k-1)$
Значение	0.0893	0.025	3.59	2.6

Где:

$N=1359$ - общее число измерений;

$K=4$ – число уровней факторов;

$D_{мг}^*$ - оценка межгрупповой дисперсии;

$D_{вг}^*$ - оценка внутригрупповой дисперсии;

f - значение критерия;

F –критическое значение из распределения Фишера для $\alpha=0,95$.

Нулевая гипотеза в отношении влияния расписания экзаменов на успеваемость групп не подтверждается ($f > F(N-1, k-1)$). Следовательно, имеет место влияния номера экзамена на успеваемость групп по результатам этого экзамена.

Выводы:

1. Использование оригинального программного стенда позволило перейти от успеваемости отдельных студентов к успеваемости учебных групп и создать условия формализации анализа влияния факторов на общее число положительных оценок, полученных по результатам сессий.
 2. Введенное понятие успеваемости групп по результатам сессий позволило корректно применить методы параметрических статистик.
 3. Применение двухуровневой декомпозиции по средней успеваемости и оценке ее дисперсии позволило предложить технологию разделения множества объектов на подмножества статистически однородных объектов. Это позволит применять параметрические статистики внутри полученных однородных подмножеств.
 4. Предложено в качестве уровня фактора использовать номер класса объектов, полученный в результате кластеризации. Такой подход позволил для достижения цели настоящего исследования корректно применять дисперсионный анализ средних успеваемости групп.
 5. Проведенный двухфакторный дисперсионный анализ доказал зависимость результатов экзаменов от классов предмета и преподавателя. Следовательно, доказано, что результаты экзамена зависят от средней успеваемости по предмету и преподавателю.
 6. Проведенный однофакторный дисперсионный анализ показал наличие примеров влияния расписания на результаты экзаменов.
-

Литература

1. Гранков М.В., Аль-Габри В.М. Постановка задачи автоматизации построения расписания экзаменов студентов вуза. Системный анализ, управление и обработка информации: Труды 5-го Международного семинара (п.Дивноморское 2-6 октября 2014 г.) Под общ. ред. Р.А.Нейдорфа: - Ростов н/Д: - ДГТУ, 2014. — 472с.
 2. Сосницкий В.Н., Потанин Н.И., Шевелева Л.В. Проблемы статистического анализа средней успеваемости студентов. Журнал «Фундаментальные исследования», №10(часть 2.), 2013, ИД «Академия естествознания». М.: 2013. - с.316-320.
 3. Сосницкий В.Н., Потанин Н.И. Вероятностный поход к анализу успеваемости студентов. Журнал «Фундаментальные исследования», №8, 2013, ИД «Академия естествознания». М.: 2014. - с.734-738.
 4. Арефьев В.П., Михальчук А.А., Филипенко Н.М. Тернарная кластеризация показателей образовательной деятельности студентов вузов. Журнал «Современные проблемы науки и образования», №3, 2014, ИД «Академия естествознания». М.: 2014. С. 2-7.
 5. Рутковский Л. Методы и технологии искусственного интеллекта / Пер.с польск. И.Д.Рудинского.-М.: Горячая линия — Телеком, 2010.- 520с.,ил.
 6. Румшинский Л.З. Математическая обработка результатов эксперимента. – М.: «Наука», 1971.-192 с.
 7. Сигел Э.Ф. Практическая бизнес-статистика.: Пер.с англ. - М.: Издательский дом «Вильямс», 2008.- 1056 с.:ил.
 8. Фишер Р.А. Статистические методы исследований.: Пер.с англ. - М.: Госстатиздат, 1957.- 266 с.
 9. Шеффе Г. Дисперсионный анализ.: Пер.с англ. - М.: «Наука», 1980.- 512с.
-

10. Балдин К.В. Теория вероятности и математическая статистика: Учебник / Башлыков В.Н., Рукоусев А.В. — 2-е изд. — М.: Издательско-торговая корпорация «Дашков и К^О», 2014. — 472 с.
11. Shiwani Rana and Roopali Garg. Application of hierarchical clustering algorithm to evaluate students performance of an institute // International Conference on Computational Intelligence & Communication Technology, no. 2 (2016). pp. 692-697.
12. Ishwank Singh, A Sai Sabitha and Abhay Bansal. Students performance analysis using clustering algorithm // International Conference - Cloud System and Big Data Engineering (Confluence), no. 6 (2016). pp. 294-299.
13. Дорохина Е.С., Хорошко А.А., Реализация программы академической и социальной адаптации студентов 1 курса в техническом ВУЗе // Инженерный вестник Дона, 2013, № 1, URL: ivdon.ru/ru/magazine/archive/n1y2013/1572/.
14. Абанина Т. И., Актуальные направления совершенствования математических знаний студентов в системе непрерывной профессиональной подготовки // Инженерный вестник Дона, 2013, № 4, URL: ivdon.ru/ru/magazine/archive/n4y2013/2112/.

Reference

1. Grankov M.V., Al'-Gabri V.M. Postanovka zadachi avtomatizatsii postroyeniya raspisaniya ekzamenov studentov vuza [Problem statement of automating the scheduling of student exams timetable at the university]. Trudy 5-go Mezhdunarodnogo seminar (p.Divnomorskoye 2-6 oktyabrya 2014 g.) Pod obshch. red. R.A.Neydorfa: Rostov n/D: DGTU, 2014. 472p.
2. Sosnitskiy V.N., Potanin N.I., Sheveleva L.V., Fundamentalnyye issledovaniya, №10 (chast' 2.), 2013, pp. 316-320.
3. Sosnitskiy V.N., Potanin N.I., Fundamentalnyye issledovaniya, №8, 2013. pp. 734-738.



4. Aref'yev V.P., Mikhal'chuk A.A., Filipenko N.M., *Sovremennyye problemy nauki i obrazovaniya*, №3, 2014. pp. 2-7.
 5. Rutkovskiy L. *Metody i tekhnologii iskusstvennogo intellekta [Methods and techniques of artificial intelligence]*. Per.s pol'sk. I.D.Rudinskogo. M.: Goryachaya liniya. Telekom, 2010. 520 p.
 6. Rumshinskiy L.Z. *Matematicheskaya obrabotka rezul'tatov eksperimenta [The mathematical processing of the experimental results]*. M.: «Nauka», 1971. 192 p.
 7. Sigel E.F. *Prakticheskaya biznes-statistika [Applied Business Statistics]*: Per.s angl. M.: Izdatel'skiy dom «Vil'yams», 2008. 1056 p.
 8. Fisher R.A. *Statisticheskiye metody issledovaniy [Statistical methods research]*. Per.s angl. M.: Gosstatizdat, 1957. 266 p.
 9. Sheffe G. *Dispersionnyy analiz [Variance analysis]*. : Per.s angl. M.: «Nauka», 1980. 512p.
 10. Baldin K.V., *Teoriya veroyatnosti i matematicheskaya statistika [Theory of probability and mathematical statistics]*: Uchebnik. Bashlykov V.N., Rukosuyev A.V. 2-ye izd. M.: Izdatel'sko-torgovaya korporatsiya «Dashkov i KO», 2014. 472p.
 11. Shiwani Rana and Roopali Garg. Application of hierarchical clustering algorithm to evaluate students' performance of an institute. *International Conference on Computational Intelligence & Communication Technology*, no. 2 (2016). pp. 692-697.
 12. Ishwank Singh, A Sai Sabitha and Abhay Bansal. Students performance analysis using clustering algorithm. *International Conference. Cloud System and Big Data Engineering (Confluence)*, no. 6 (2016). pp. 294-299.
 13. Dorokhina Ye.S., Khoroshko A.A., *Inzhenernyy vestnik Dona (Rus)*, 2013, № 1 URL: ivdon.ru/ru/magazine/archive/n1y2013/1572/.
 14. Abanina T. I., *Inzhenernyy vestnik Dona (Rus)*, 2013, № 4 URL: ivdon.ru/ru/magazine/archive/n4y2013/2112/.
-