

УДК 316.6

МАШИННОЕ ОБУЧЕНИЕ МОДЕЛИ ИНФОРМАЦИОННОЙ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ПО ВОПРОСАМ ИНДИВИДУАЛИЗАЦИИ ОБРАЗОВАНИЯ

Таратухина Юлия Викторовна,

*канд. филол. наук, доцент факультета бизнеса и менеджмента,
e-mail: jtaratuhina@hse.ru,*

Национальный исследовательский университет «Высшая школа экономики», г. Москва,

Барт Татьяна Вячеславовна,

*канд. экон. наук, доцент кафедры менеджмента и маркетинга,
e-mail: tbart@mail.ru,*

Московский университет им С.Ю.Витте, г. Москва,

Власов Владимир Владимирович,

*магистрант,
e-mail:vladimir.vlasov@mail.ru,*

Национальный исследовательский университет «Высшая школа экономики», г. Москва

Обучение модели информационной рекомендательной системы связано с исследованием применяемых математических и информационных методов и моделей, их комбинаций с целью обеспечения необходимой точности формируемых прогнозов и выводов. В статье рассматривается машинное обучение модели рекомендательной системы с применением статистических методов и анализа больших данных, направленной на решение вопросов индивидуализации образования. В данном случае точность машинного обучения модели зависит от типа статистической модели, используемой для прогнозирования вероятности возникновения некоторого события по значениям множества признаков, а также обучающей выборки, используемой для подбора параметров, и функции регуляризации, используемой для улучшения обобщающей способности получающейся модели. В рамках исследования проверяются модели на основе логистической регрессии, методах наивного байесовского классификатора (Naïve Bayes), регрессии типа Лассо. Экспериментально подтверждается теоретическое предположение о возможности создания рекомендательной системы по вопросам индивидуализации образования на основе массива образовательных данных, включающего результаты учебной и внеучебной деятельности учащихся. Формулируются выводы о наличии корреляционных зависимостей в данных, которые могут быть использованы для повышения точности обучения модели рекомендательной системы.

Ключевые слова: индивидуализация образования, машинное обучение модели, анализ больших данных, рекомендательные системы

MACHINE LEARNING MODELS OF INFORMATION RECOMMENDATION SYSTEM ON INDIVIDUALIZATION OF EDUCATION

Taratukhina Yu.V.,

*Associate Professor, department of innovation and business in information technologies,
e-mail: jtaratuhina@hse.ru,*

National Research University Higher School of Economics, Moscow,

Bart T.V.,

*Ph.D., Assistant Professor, Assistant Professor of the management and marketing chair,
e-mail: tbart@mail.ru,*

Moscow Witte University, Moscow,

Vlasov V.V.,

*master,
e-mail:vladimir.vlasov@mail.ru,*

National Research University Higher School of Economics, Moscow

Training model information recommendation system is associated with the study of applied mathematical and information methods and models, their combinations in order to ensure the necessary accuracy of the forecasts and conclusions. The article deals machine learning of model recommendation system using statistical methods and analysis of big data, aimed at addressing the issues of individualization of education. In this case, the accuracy of the machine learning model depends on the type of statistical model used to predict the probability of some event from the values of the set of features, as well as the training sample used to select the parameters, and the regularization function used to improve the generalizing ability of the resulting model. The study tested models based on logistic regression, methods of naive Bayesian classifier (Naïve Bayes), lasso-type regression. Experimentally confirmed the theoretical assumption about the possibility of creating a recommendation system on the individualization of education on the basis of an array of educational data, including the results of educational and extracurricular activities of students. Conclusions about the presence of correlation dependencies in the data, which can be used to improve the accuracy of the model of the recommendation system, are formulated.

Keywords: individualization of education, machine learning models, big data analysis, recommendation systems

DOI 10.21777/2500-2112-2019-2-7-14

Введение

На сегодняшний день основной задачей системы образования становится подготовка индивида к жизни в быстроменяющемся мире, в глобальном поликультурном пространстве. По сути, образовательное пространство представляет собой сосуществование различных образовательных систем и моделей, в основе которых лежат дифференцированные культурные, мировоззренческие, религиозные, философские, ценностные картины мира. Кроме интеграционных процессов происходит трансформация форм образования. Наряду с традиционной классической моделью создаются и внедряются в практику инновационные формы обучения, основанные на современных информационно-коммуникационных технологиях (ИКТ). К ним относятся массовые открытые онлайн-курсы, в том числе с применением предметно-языкового интегрированного подхода, бесплатные онлайн курсы от ведущих университетов и школ мира и др. В этой связи можно отметить ряд следующих популярных образовательных платформ: www.coursera.org (Университеты Стенфорда и Принстона (США), университеты Пекина, Гонконга, Торонто, Тель-Авива); www.edx.org (Университет Беркли, Гарварда, Массачусетский технологический институт); <http://netology.ru> – Нетология, Национальная платформа открытого образования; <https://www.lektorium.tv> – Лекториум; <http://universarium.org> – Универсариум.

Актуальными являются технологии обучения, основанные на виртуальном взаимодействии (Second Life), а также формы обучения, основанные на геймификации – играх, тренингах, симуляторах и т.п., в форме которых подается образовательный контент (www.edutainme.ru). Распространенным явлением в настоящее время является мобильное обучение, использование технологий дополненной реальности, машинное обучение, использование интеллектуальных тьюторских систем. По мере развития форм электронного обучения, в том числе с применением дистанционных образовательных технологий, все больше внимания уделяется вопросам персонализации образовательной среды обучающегося, отслеживанию и моделированию его индивидуальной образовательной траектории с целью повышения качества и эффективности обучения [3]. В этом контексте роль преподавателя меняется – он становится не только ретранслятором знаний, но и опытным наставником и сопровождающим лицом, тьютором, помогающим выстраивать обучающемуся его индивидуальную образовательную траекторию. В идеале, существование индивида в информационном обществе подразумевает самостоятельную «добычу знаний» и «управление» ими и, как следствие, высокую роль самостоятельности и ответственности. Формат непрерывного обучения (life long Learning; education throw life) должен реализовываться именно в контексте, к которому побуждает информационная среда: непрерывность, активное использование ИКТ, автодидактичность, использование доступных глобальных сетевых ресурсов. Одними из таких возможностей являются реализация концепции blended learning (смешанное обучение) и анализ деятельности обучающихся с применением технологий Big Data [1, 6, 7].

Вопросы анализа образовательных данных

Анализ результатов успеваемости и достижений обучающихся является необходимым инструментом для индивидуализации обучения. Кроме того он позволяет решить множество задач при внедрении в учебные заведения любого уровня. К этим задачам можно отнести кластеризацию обучающихся и педагогов для выявления зависимостей и связей между ними, оценку качества образования, оценку риска получения обучающимися неудовлетворительной оценки, автоматическое построение рекомендаций по использованию информационных ресурсов и материалов для более эффективного освоения образовательной программы и др. Для многих учащихся в школе одним из важных вопросов является выбор дальнейшей траектории обучения и в данном вопросе информационные рекомендательные системы в связке с выбором формальных онлайн и смешанных курсов могут оказать положительное влияние на ситуацию [4]. Настроенная рекомендательная система может позволить не только проходить курсы онлайн, но и будет советовать, какие из них стоит пройти тому или иному учащемуся, что позволит качественно работать с системой и персонализировать образование.

При обучении модели рекомендательной системы возникает ряд вопросов о возможности ее обучения, в том числе: можно ли обеспечить необходимую точность обучения на основании записей об оценках и внеучебной деятельности обучающихся в рамках имеющегося массива данных системы и других открытых информационных ресурсов; будут ли предсказания системы ценными для обучающихся; можно ли создать эталонный цифровой портрет обучающегося, который хочет работать с той или иной сферой знаний, и учитывать индивидуальные качества и предпочтения обучающихся при формировании индивидуальной образовательной траектории; какая из полученных моделей системы наиболее точно делает прогноз и выдает рекомендации. Для получения ответов на данные вопросы в рамках исследования должна быть создана и проверена тестовая версия подобной системы, основанной на машинном обучении модели.

В зарубежных периодических изданиях опубликовано достаточно много научных статей, посвященных анализу образовательных данных и созданию рекомендательных систем в области образования [8, 9]. В Российской Федерации данная тема в настоящее время также становится актуальной. Основываясь на анализе зарубежной и отечественной научной и специальной литературы по данной теме, можно сказать, что рекомендации в формальном образовании практически не развиты, в качестве объекта рассматриваются отдельные медиа-ресурсы, а не курсы в целом. Рекомендательные системы в образовании развиты только в массовых открытых онлайн курсах на крупных ресурсах, таких как eDX или Coursera. При этом они основываются не на данных об успеваемости и внеучебной деятельности обучающихся и настроены на другие типы запросов.

Данная работа ставит своей задачей исследование именно российских образовательных данных, которые могут отличаться ввиду культурных и социальных факторов, а также возможности применения рекомендательной системы в рамках формального образования.

Современный образовательный процесс генерирует большой объем данных, которые могут быть использованы для проведения исследований, посвященных различным аспектам образовательной деятельности. Данный факт достаточно очевиден при использовании образовательных онлайн платформ, однако и в традиционной системе обучения при непосредственном взаимодействии преподавателей и обучающихся создается достаточный объем данных для проведения анализа с применением методов машинного обучения. Накопленные данные являются крайне важными для анализа процесса обучения с целью улучшения и развития образовательного процесса [9].

В настоящее время в Российской Федерации реализуется множество проектов по внедрению ИКТ в образование, но для того чтобы они работали с полной отдачей необходимо создать систему автоматической и, что более важно, постоянной оценки процесса обучения. Для этого необходимо наблюдать за различными его аспектами, включая результаты учебной и внеучебной деятельности обучающихся, уровень взаимодействия преподавателей и обучающихся, направление использования мультимедийных ресурсов, качество управления учебным процессом. При внедрении новых технологий увеличивается и объем данных, которые могут быть собраны в рамках образовательного процесса для последующего анализа, что, с одной стороны, делает задачу машинного обучения все более важной, с другой

стороны, улучшает качество моделей, которые могут быть при помощи подобного анализа построены.

Вопросы анализа образовательных данных исследуются учеными многих стран. Работы, посвященные данной теме, можно разделить на четыре группы: оценка эффективности электронных средств при традиционной концепции обучения, анализ действий преподавателей и обучающихся, выявление групп риска среди обучающихся, исследование работы с различными медийными источниками. Ряд работ посвящается оценке того как влияют ИКТ на образование в целом. Примером этого можно назвать семантический анализ форумов и блогов, посвященных образовательным платформам [1]. В данном случае исследователи при помощи трех различных методов семантического анализа (Information Gain, Mutual Information, CHI statistics) создали специальную модель, которая определяет является ли данная запись позитивно или негативно окрашенной. Модель может быть полезна для оценки того, как пользователи реагируют на новую технологию в образовании и упростить оценку ее качества. Также к этому классу исследований можно отнести те, целью которых является разделение обучающихся на различные группы в зависимости от того, как быстро и качественно они могут осваивать курсы. При этом применяются различные методы, такие как, нейронные сети, Naïve Bayes, SMO и др., которые также позволяют выделить основные факторы, влияющие на эффективность освоения курсов обучающимися.

Сбор и обработка экспериментального массива данных

Исследование образовательных данных включает следующие этапы:

1. Сбор данных об успеваемости обучающихся и их внеучебных достижениях.
2. Обработка полученного массива данных и перевод переменных в категориальный и числовой вид [2].
3. Создание моделей классификации «один-против-всех» [5].
4. Выбор наиболее качественных видов моделей для решения поставленной задачи.
5. Обучение модели рекомендательной системы на основе анализа образовательных данных [7].
6. Оценка коэффициентов модели с применением метода Лассо-регрессии [12].

Для проведения эксперимента по обучению модели использовался массив образовательных данных, созданный по результатам анонимного опроса выпускников школы в сети Интернет. Цель опроса – выявить, какие предметы выбрали бы учащиеся, если бы обучались в online-системе, их оценки в школе по основным предметам, участие в олимпиадах и других видах внеучебной деятельности. Результаты опроса были преобразованы в массив данных, содержащий числовые и бинарные показатели, на основании которых можно сделать прогнозы о том, какие предметы учащиеся выбрали бы для изучения. В завершение работы была построена и проверена рекомендательная модель.

Опрос состоит из трех частей (таблица 1): на какую специальность или направление респондент хочет поступить или поступил и какие предметы он бы выбрал для онлайн изучения на базовом или продвинутом уровне; оценки учащегося по основным предметам; участие в олимпиадах различного уровня и внеклассных мероприятиях.

Таблица 1 – Структура опроса

Части опроса	Список вопросов
1	«Я поступил (или хотел бы поступить) на направление связанное с...» «Если бы я мог проходить на базовой основе школьный предмет в системе онлайн и при этом отказаться от обучения по нему в традиционной школе, то это были бы следующие предметы...» «Если бы я мог проходить на продвинутой основе (частично с вузовской программой) школьный предмет в системе онлайн и при этом отказаться от обучения по нему в традиционной школе, то это были бы следующие предметы...» «Основная причина почему я стал бы учиться в системе онлайн»
2	12 типовых вопросов про уровень оценок в школе
3	3 вопроса об участии в окружном, областном и финальных этапах Всероссийской олимпиады школьников по разным предметам Вопросы об участии в олимпиадах, проводимых вузами Вопросы о достижениях во внеучебной деятельности

Первая часть опроса служит для создания переменных, которые с одной стороны важны для построения прогноза, с другой стороны являются теми целевыми переменными, которые должны быть предсказаны системой для каждого отдельного пользователя. Первый вопрос ставит своей целью узнать дальнейшее направление обучения респондента. Полученные из него переменные являются целевыми для системы в целом. Второй и третий вопросы являются более узкими и предлагают составить свой список предметов, которые респондент хотел бы проходить на базовом или продвинутом уровне. Оба эти списка преобразуются в бинарные категориальные переменные для последующего анализа [2]. Вторая часть опроса служит для того, чтобы узнать какие оценки были у респондента по предметам в школе. Ответы варьируются от неудовлетворительных оценок до отличных. Список предметов (алгебра, геометрия, русский, литература, история, иностранный язык, физика, химия, география, обществознание, биология, информатика) соответствует тому, какие предметы предлагаются для изучения в системе онлайн на базовом и продвинутом уровне. Третья часть опроса создана для того, чтобы ознакомиться с внеучебными и олимпиадными успехами респондента. Здесь список предметов несколько меняется: алгебра и геометрия объединены в математику. В качестве внеучебной деятельности предполагаются участие в спортивных и общественных мероприятиях, включая волонтерскую деятельность, и др.

Полученный массив данных перерабатывается компьютерной программой в формат, который подходит для построения математических моделей. В рамках исследования такая программа была реализована на языке Python с использованием библиотек для анализа данных *pandas* и *random* и программной среды *Jupyter Notebook*. В реализованной программе стоит отметить два важных момента. Во-первых, ответы пользователей представлены категориальными бинарными переменными, что позволяет использовать их для создания прогнозов. Во-вторых, ответы пользователей по вопросам успеваемости представлены как числовые переменные, примерно равные ожидаемому среднему баллу по предмету. Так, в случае, если респондент отвечал на вопрос об оценке «только пятерки», программа переводит этот ответ в случайное значение от 4,7 до 5,0. Это сделано для большей схожести с теми данными, которые могут собираться в автоматическом режиме. Всего экспериментальный массив данных содержит ответы 252 респондентов, каждому из них соответствует 96 категориальных и числовых переменных, которые могут использоваться для составления предсказаний.

Рассмотрим набор данных при помощи стандартных статистических показателей. Перед тем, как перейти к описанию самих переменных, проведем анализ как респонденты отвечали на вопрос «какова основная причина, из-за которой они выбрали бы онлайн курсы вместо обычных школьных». Лишь 16 % опрошенных посчитали, что они не стали бы учиться в системе онлайн. Около 40 % респондентов связали желание учиться в онлайн системе с недостаточным качеством преподавания и/или необходимостью расширения возможностей для подготовки к ЕГЭ. Больше 30 % респондентов говорят о том, что система онлайн позволила бы им оставить больше свободного времени, которое они могли бы направить на другую деятельность. 10 % респондентов говорят о том, что для них наиболее важной является возможность изменить свой распорядок дня.

Теперь перейдем к описанию того, какими получились значения числовых и категориальных переменных. Первой группой таких переменных являются оценки. Все средние значения находятся в диапазоне от 4 до 4,7 баллов. Судя по данной выборке наилучшие оценки школьники получают по литературе, иностранному и русскому языкам (около 4,6 баллов); наихудшие по химии и физике (около 4,15 баллов) и алгебре и геометрии (4,27 балла). Стоит заметить, что этим предметам соответствует и наибольшая дисперсия, что может свидетельствовать о том, что в отличие от других эти предметы либо по разному преподаются в разных школах, либо наиболее по разному даются учащимся. Далее переходим к первому типу категориальных переменных – поступление в вузы. Больше всего (36 %) опрошенных выбрали гуманитарные направления (специальности), на втором месте экономические (29 %), меньше всего респондентов выбрали педагогические, медицинские и естественнонаучные (около 8 % каждый) направления подготовки. Все предметы, доступные для выбора, можно было выбрать как в базовом, так и в продвинутом варианте. Среди базовых предметов наиболее популярными оказались география (44 %), информатика (38 %), история (31 %), биология, химия и физика (около 27 %). Наименее популярными базовыми предметами оказались русский и иностранный языки (около 10 %). Предметы продвинутого уровня изучения распределились следующим образом: на первом месте иностранный

язык (46 %), на втором месте алгебра (28 %); меньше всего у географии (7 %), биологии и химии (по 10 %); остальные предметы находятся в районе 15–20 %. Изучением предметов на базовом уровне среди опрошенных хотят заниматься несколько чаще, чем на продвинутом уровне (83 % против 78 %). Следует заметить, что среди опрошенных очень высокий процент учащихся, которые принимали участие в олимпиадах, в том числе, 12 % участвовали в финальном этапе Всероссийской школьной олимпиады. Наиболее популярными являются олимпиады по иностранному языку, русскому, математике, литературе и общественному. 17 % респондентов занимались волонтерской деятельностью, 43 % спортом, 64 % – общественной деятельностью. Анализ полученных данных был проведен с применением дополнительных библиотек sklearn, numpy, math и scipy.

На основе подготовленного массива данных были применены и протестированы следующие модели рекомендательной системы:

- логистическая регрессия [11];
- мультиномиальная и многомерная модели на основе наивного байесовского классификатора (Naïve Bayes) [10];
- регрессия типа Лассо или L1-регуляризация (для оценки значимости переменных) [12].

Из них для дальнейшего использования были выбраны логистическая регрессия, мультиномиальная и многомерная модели (Бернулевский Naïve Bayes). Во-первых, все три эти модели показали себя практически одинаково эффективными при подсчете очков (функция AUC-score на языке Python). Во-вторых они подходят для дальнейших предсказаний, так как обладают возможностью показывать не только предсказание, но и его вероятность.

Полученная система должна была выдать k предсказаний для пользователя. Попытка выдать предсказание считалась успешной, если среди этих предсказаний было хотя бы одно верное. Данный результат сравнивался с тем, как при данном k вела себя случайная модель. Для оценки предсказаний доверительный интервал был построен с $p\text{-value}=0.001$. Результаты проверки моделей представлены в таблице 2.

Таблица 2 – Оценка предсказания предпочтений учащегося по выбору курсов для онлайн обучения

k	Доверительный интервал	Логистическая регрессия	Multinomial_NB	Bernoulli_NB
1	[0.14, 0.41]	0.736	0.66	0.601
2	[0.308, 0.591]	0.881	0.786	0.786
3	[0.431, 0.707]	0.894	0.883	0.854
4	[0.512, 0.79]	0.934	0.922	0.902
5	[0.591, 0.841]	0.974	0.941	0.941

Как можно видеть из данной таблицы все три исследуемые модели постоянно находятся выше доверительного интервала случайной модели, что свидетельствует о том, что все они являются значимыми с достаточно большой вероятностью. При этом и другие показатели моделей говорят о том, что они могут быть использованы для создания рекомендательной системы. Так в среднем при $k=5$ более чем 2,5 прогноза являются точными, что позволяет утверждать, что пользователь с достаточно высокой вероятностью может воспользоваться системой для получения рекомендаций.

Другая рекомендательная модель была построена для выявления того на какую специальность человек поступил или хотел бы поступить в дальнейшем. Для нее доверительный интервал был построен с $p\text{-value}=0.01$. Так как всего предполагалось 7 вариантов ответа на данный вопрос, были протестированы только первые четыре варианта модели (от одного до четырех лучших). Результаты проведенного исследования показаны в таблице 3.

Таблица 3 – Оценка предсказания предпочтений учащегося по выбору будущей специальности

k	Доверительный интервал	Логистическая регрессия	Multinomial_NB	Bernoulli_NB
1	[0.161, 0.416]	0.428	0.523	0.492
2	[0.339, 0.602]	0.662	0.711	0.656
3	[0.466, 0.723]	0.753	0.82	0.796
4	[0.564, 0.801]	0.87	0.875	0.859

Стоит отметить, что, несмотря на то, что модель является статистически значимой при выбранном p -value, она показала себя несколько хуже, чем первая, построенная для предсказания выбираемых предметов для онлайн обучения. Это может быть связано с тем, что из-за особенностей опроса в одной группе было объединено слишком большое количество специальностей, которые достаточно сильно различаются. Однако, несмотря на это рекомендательная система все же имеет возможность выводить в списке первых трех рекомендаций верные направления для около 80 % учащихся.

Также был проведен анализ важности показателей для модели. Для этого была создана регрессия типа Лассо, которая выгодна тем, что обнуляет наименее важные для предсказания коэффициенты. Важно отметить, что сравнивать здесь имеет смысл категориальные переменные отдельно от численных из-за разницы в «шкале измерений». Необходимо отметить, что для более точного анализа стоит рассматривать модель для каждого предмета по отдельности. Например, на продвинутой алгебре идут люди, которые:

- также идут на продвинутую геометрию, иностранный язык, обществознание и физику;
- также идут на базовые химию и физику, но не идут на базовые обществознание и геометрию;
- не участвовали в олимпиадах, в первую очередь по русскому и литературе;
- имеют более высокие оценки по геометрии и физике, более низкие по информатике.

В целом для развития подобной модели важным является получить больше данных для обучения, однако даже на таком уровне можно заметить, что выбор предметов зависит от оценок, выбора других предметов и олимпиад.

Заключение

Исследование подтвердило выдвинутую гипотезу о возможности создания рекомендательной системы по вопросам индивидуализации образования с использованием данных об успеваемости и внеучебной деятельности обучающихся. На примере результатов обработки экспериментального массива данных демонстрируется достигнутая точность формируемых системой рекомендаций с применением разных методов обучения модели и для разных видов запросов пользователей. Следует заметить, что данное исследование обладает рядом ограничений, которые могут быть решены автоматическим сбором данных. Это позволит, с одной стороны, улучшить качество получаемых данных, с другой, увеличить количество людей, по данным которых можно построить модель. Из-за ограниченности массива данных (только 250 примеров) не было возможности четко оценить границы классов в моделях «один против остальных» из-за опасности переобучения модели. Несмотря на это достаточно важным является именно подтверждение выбранной гипотезы, которое говорит о целесообразности проведения дополнительных исследований и создании рекомендательных систем для улучшения и персонализации образования. Реализация наилучшей модели рекомендательной системы позволит автоматически создавать эталонный цифровой портрет обучающегося, который намерен работать с той или иной сферой знаний, а также учитывать профессиональные и личностные запросы конкретного обучающегося при формировании индивидуальной образовательной траектории.

Список литературы

1. Белоножко, П.П. Анализ образовательных данных: направления и перспективы применения / П.П. Белоножко, А.П. Карпенко, Д.А. Храмов // Интернет-журнал «Науковедение». – Том 9. – № 4 (2017) [Электронный ресурс]. Режим доступа: <http://naukovedenie.ru/PDF/15TVN417.pdf> (дата обращения: 10.04.2019).
2. Боровиков В.П. STATISTICA: искусство анализа данных на компьютере. (Для профессионалов). – СПб.: Питер, 2001.
3. Гладкая, Е.С. Технологии тьюторского сопровождения [Текст]: учебное пособие / Е.С. Гладкая, З.И. Тюмасева. – Челябинск: Изд-во Юж.-Урал. гос. гуман.-пед. ун-та, 2017. – 93 с.
4. Гомзин, А.Г. Системы рекомендаций: обзор современных подходов / А.Г. Гомзин, А.В. Коршунов // Труды Института системного программирования РАН. – 2012. – Т. 22.

5. Решетова Д.Г., Максимов Ю.В. Сложность многоклассового классификатора один-против-всех // Труды МФТИ.– 2015. – Том 7. – № 4. – С. 59–65.
6. Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance / Li, Liang-Yi; Tsai, Chin-Chung // Computers & Education. – 2017. – Volume. 114.
7. A Method of Unstructured Information Process in Computer Teaching Evaluation System Based on Data Mining Technology / Quan Liu; Yongjun Peng // 2013 International Conference on Communication Systems and Network Technologies (CSNT), Gwalior, India, 2013.
8. Classification and prediction based data mining algorithms to predict slow learners in education sector / P. Kaur; M. Singh; G. S. Josan // Procedia Computer Science. – 2015. – Volume. 57. – P. 500–508.
9. Developing early warning systems to predict students' online learning performance / Ya-Han Hu; Chia-Lun Lo; Sheng-Pao Shih // Computers in Human Behavior. – 2014. – Volume 36. – P. 469–480.
10. Domingos, Pedro & Michael Pazzani (1997) «On the optimality of the simple Bayesian classifier under zero-one loss». Machine Learning, 29:103-137.
11. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2001. 533 p.
12. Tibshirani R. Regression shrinkage and Selection via the Lasso //Journal of the Royal Statistical Society. Series B (Methodological). – 1996. – Vol. 32. – № 1. – P. 267–288.

References

1. Belonozhko, P.P. Analiz obrazovatel'nyh dannyh: napravleniya i perspektivy primeneniya / P.P. Belonozhko, A.P. Karpenko, D.A. Hramov // Internet-zhurnal «Naukovedenie». – Tom 9. – № 4 (2017). – URL: <http://naukovedenie.ru/PDF/15TVN417.pdf> (data obrashcheniya: 10.04.2019).
2. Borovikov V.P. STATISTICA: iskusstvo analiza dannyh na komp'yutere. (Dlya professionalov). – SPb.: Piter, 2001.
3. Gladkaya, E.S. Tekhnologii t'torskogo soprovozhdeniya [Tekst]: uchebnoe posobie / E.S. Gladkaya, Z.I. Tyumaseva. – Chelyabinsk: Izd-vo Yuzh.-Ural. gos. guman.-ped. un-ta, 2017. – 93 s.
4. Gomzin, A.G. Sistemy rekomendacij: obzor sovremennyh podhodov / A.G. Gomzin, A.V. Korshunov // Trudy Instituta sistemnogo programmirovaniya RAN. – 2012. – T. 22.
5. Reshetova D.G., Maksimov Yu.V. Slozhnost' mnogoklassovogo klassifikatora odin-protiv-vsekh // Trudy MFTI, 2015. – Tom 7. – № 4. – S. 59–65.
6. Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance / Li, Liang-Yi; Tsai, Chin-Chung // Computers & Education. – 2017. – Volume. 114.
7. A Method of Unstructured Information Process in Computer Teaching Evaluation System Based on Data Mining Technology / Quan Liu; Yongjun Peng // 2013 International Conference on Communication Systems and Network Technologies (CSNT), Gwalior, India, 2013.
8. Classification and prediction based data mining algorithms to predict slow learners in education sector / P. Kaur; M. Singh; G.S. Josan // Procedia Computer Science. – 2015. – Volume. 57. – P. 500–508.
9. Developing early warning systems to predict students' online learning performance / Ya-Han Hu; Chia-Lun Lo; Sheng-Pao Shih // Computers in Human Behavior. – 2014. – Volume 36. – P. 469–480.
10. Domingos, Pedro & Michael Pazzani (1997) «On the optimality of the simple Bayesian classifier under zero-one loss». Machine Learning, 29:103–137.
11. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2001. – 533 p.
12. Tibshirani R. Regression shrinkage and Selection via the Lasso //Journal of the Royal Statistical Society. Series B (Methodological). – 1996. – Vol. 32. – № 1. – P. 267–288.