

МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ГРУППЫ РИСКА СТУДЕНТОВ

ARTIFICIAL INTELLIGENCE METHODS IN THE PROBLEM OF FORECASTING STUDENTS RISK GROUP

Сергей Владимирович Русаков

Sergei Vladimirovich Rusakov

доктор физико-математических наук,
профессор, заведующий кафедрой
прикладной математики и информатики
rusakov@psu.ru

Ольга Леонидовна Русакова

Olga Leonidovna Rusakova

кандидат физико-математических наук, доцент,
доцент кафедры прикладной
математики и информатики
rol58@yandex.ru

Наталья Николаевна Накарякова

Natalia Nikolaevna Nakariakova

магистр
nata_nakar@mail.ru

Елена Станиславовна Чингаева

Elena Stsnislavivna Chingaeva

бакалавр
chingaeva.elena@gmail.com

ФГБОУ ВО «Пермский государственный
национальный исследовательский
университет», Пермь, Россия

Perm State National Research University, Perm,
Russia

Аннотация. Описана методика классификации студентов с целью определения тех из них (группы риска), которые будут отчислены уже по итогам первого триместра обучения. При этом используются данные о студентах, имеющиеся в вузе на начало учебы. В качестве инструментов классификации применяются нейронные сети и деревья решений.

Abstract. The paper describes the methodology of classifying students in order to determine those of them (risk groups) that will be deducted already according to the results of the first trimester of study. At the same time, data on students available at the university at the beginning of studies are used. Neural networks and decision trees are used as classification tools.

Ключевые слова: классификация, нейронные сети, деревья решений, рейтинговые баллы, группа риска.

Keywords: classification, neural networks, decision trees, rating points, risk group.

Проблема определения группы риска на основе анализа результатов ЕГЭ, т. е. выявления студентов, которые могут быть отчислены в самом начале обучения, является весьма актуальной. Как известно, количество обучающихся напрямую связано с финансированием вуза, с планированием бюджетных средств на следующий год. Связи результатов ЕГЭ с успешностью обучения посвящено достаточно много исследований, причем чаще всего в качестве математического аппарата используется корреляционный и регрессионный анализ, при этом результаты исследований весьма противоречивы. Так, например, в работе Т. Е. Хавенсон и А. А. Соловьевой [1] была проверена статистическая гипотеза, в равной ли степени эффективны баллы ЕГЭ по разным предметам в качестве прогностического фактора дальнейшей успеваемости. Результаты анализа показали, что предсказательная способность баллов ЕГЭ по отдельным предметам примерно одинакова, но все же ЕГЭ по математике и русскому языку являются лучшими предикторами для подавляющего большинства направлений. В то же время в работе В. И. Переяславского [2] были рассчитаны коэффициенты корреляции между результатами ЕГЭ по математике и русскому языку и успеваемостью студентов. Результаты исследования показали низкую связь между баллами ЕГЭ по математике и успеваемостью в школе и в вузе и более высокую связь результатов экзамена по русскому языку и успеваемостью по математике в вузе.

Очевидно, что результаты ЕГЭ не всегда характеризуют полноценный уровень обученности будущих студентов. Существенное влияние на этот уровень оказывают различные процедуры «натаскивания» (в самой школе, на подготовительных курсах, с помощью репетиторов и т. п.). Более объективной характеристикой является качество обучения в конкретном образовательном учреждении. Так, в работе Е. А. Поповой и М. В. Шеиной [3] дополнительно рассматривалось влияние характеристики школ, в частности, специализации школы, на качество школьного образования. В ходе исследования с помощью эконометрического моделирования было доказано, что студенты из школ с высоким средним баллом ЕГЭ по мате-

матике учатся в среднем лучше, чем студенты из школ с низким или средним его значением.

Очевидно, что на успешность в учебе студентов-первокурсников влияют и такие материальные факторы, как уровень благосостояния семьи и бытовые условия в период обучения. В работе К. Ю. Татусь и С. В. Кузьминой [4] обобщены исследования взаимосвязи динамики успеваемости студентов в различных типах родительских семей и сделаны выводы, что успеваемость студентов из малодетных семей значительно выше, чем многодетных. Среди студентов из малообеспеченных семей обучаются хорошо и удовлетворительно 85 %, отлично — 5 % и учатся с долгами — 10 %. В статье Е. П. Богданова и А. В. Суханова [5] приведены результаты исследования успеваемости студентов первого курса с учетом довузовского обучения, места проживания во время учебы в вузе и социального статуса родителей. Авторы статьи, анализируя влияние социального статуса родителей на успеваемость, выборку данных поделили на пять групп: предприниматель, пенсионер, фермер, рабочий, госслужащий. В ходе исследования авторы установили, что лучшую успеваемость имеют дети госслужащих, рабочих, фермеров, пенсионеров, а наихудшую — дети предпринимателей. Авторы считают, что такие результаты объясняются тем, что студенты из семей предпринимателей не стеснены в материальных благах и, возможно, меньше мотивированы к обучению. В то же время, студенты из малообеспеченных семей наиболее мотивированы к учебе с целью получить престижную профессию и улучшить материальное положение. Успехи студентов из семей госслужащих, возможно, связаны с тем, что родители могут больше уделять внимание детям из-за четкого регламента рабочего дня.

В работах авторов настоящей статьи («Нейросетевая модель прогнозирования группы риска по успеваемости» [6, 7]) для определения группы риска использовались нейросетевая модель и дерево решений, полученные на периоде наборов студентов 2014, 2015, 2016 гг. Эти исследования показали, что предлагаемые методы позволяют решить поставленную задачу с приемлемой точностью. В настоящей работе эти два метода были применены для одного и того

Факт	Дерево решений		Нейронная сеть	
	Прогноз «отчислен»	Прогноз «не отчислен»	Прогноз «отчислен»	Прогноз «не отчислен»
Отчислен	16	1	8	9
Не отчислен	32	52	39	45

же набора данных, что позволило сопоставить и оценить применимость этих двух методов.

В предлагаемом исследовании использовались данные о студентах направления прикладная математика и информатика Пермского государственного национального исследовательского университета (ПГНИУ). Для обучения (нейронные сети и деревья решений) использовался период набора студентов за 4 года: 2014–2017 гг. Это время характеризовалось постоянством количества бюджетных мест (85 человек), одной рабочей программой и стабильным составом преподавателей, работающих на первом курсе. Наличие в программе таких непростых базовых дисциплин, как «Математический анализ», «Алгебра и аналитическая геометрия», «Алгоритмизация и программирование» приводит к тому, что уже по итогам первого триместра из-за академической неуспеваемости отчисляются от 15 до 25 % студентов-первокурсников.

При обучении использовались интеллектуальные системы классификации и каждый студент характеризовался по следующим параметрам:

- баллы ЕГЭ по информатике, математике и русскому языку;
- учебное заведение, в котором получено среднее общее образование (1 — специализированная школа, лицей, гимназия г. Перми; 2 — специализированная школа, лицей, гимназия населенного пункта Пермского края; 3 — СОШ г. Перми; 4 — СОШ населенного пункта Пермского края);
- место проживания студента во время обучения в вузе (дом, общежитие, съемная квартира);
- иностранный язык, изучаемый в школе (английский, другой иностранный язык).

Для построения нейросетевой модели использовалась нейронная сеть с одним скрытым слоем, содержащим 5–7 нейронов. Построенное дерево решений характеризовалось следу-

ющими параметрами: выполнена балансировка классов, максимальная глубина дерева — 6, минимальное количество экземпляров в листе — 3.

Построенные модели были использованы для прогнозирования группы риска среди студентов набора 2018 г., когда было уже 100 бюджетных мест. По итогам первого триместра были отчислены 17 человек. Результаты расчетов представлены в таблице.

Группу риска составляют те студенты, для которых прогноз имеет значение «отчислен». Таким образом, видно, что число студентов, попавших в эту группу, для обеих моделей практически совпадает и она содержит список студентов с большим запасом (примерно в 3 раза больше, чем было реально отчислено). Но вот ошибку первого рода у дерева решений составляет только один человек, в то время как у нейронной сети — 9.

Дерево решений позволяет построить следующие решающие правила:

- студент с вероятностью 100 % попадает в группу риска, если набранное количество баллов ЕГЭ по информатике менее 51;
- студент, поступивший в университет из другого населенного пункта, с вероятностью 100 % попадает в группу риска, если набранное количество баллов ЕГЭ по информатике менее 75, по математике менее 61, по русскому языку менее 96;
- студент с вероятностью 100 % попадает в группу риска, если набранное количество баллов ЕГЭ по математике менее 70 и по русскому языку менее 68, несмотря на высокие баллы ЕГЭ по информатике.

В первую пятерку по значимости параметров классификации вошли следующие (в порядке убывания значимости): баллы ЕГЭ по информатике (32 %); баллы ЕГЭ по математике (28 %); баллы ЕГЭ по русскому языку (23 %); учебное заведение (школа), которое закончил студент (8 %); место проживания студента в период учебы (5 %). Таким образом, видно, что

наиболее значимыми для успешности обучения на первом этапе являются результаты ЕГЭ, но их одних недостаточно для построения удовлетворительного прогноза.

В результате исследования можно констатировать, что более точный результат прогнозирования дает модель в виде дерева решений, при этом группа риска определяется с большой избыточностью (почти половина набора).

Также в ходе исследования была получена оценка значимости факторов, влияющих на попадание студентов в группу риска, и было построено решающее правило «вычисления аутсайдеров».

Информация о студентах, попавших в группу риска, может быть полезной для тьюторов (кураторов) и преподавателей, ведущих занятия на первом курсе.

Список литературы

1. Хавенсон, Т. Е. Связь результатов Единого государственного экзамена и успеваемости в вузе / Т. Е. Хавенсон, А. А. Соловьева // Вопросы образования. 2014. № 1. С. 186–199.

2. Переяславский, В. И. Исследование корреляции между результатами ЕГЭ по математике абитуриентов и их успеваемостью в вузе / В. И. Переяславский, Л. Б. Переяславская. Текст: непосредственный // Вестник ассоциации вузов туризма и сервиса. 2014. Т. 8, № 4. С. 49–53.

3. Попова, Е. А. Успеваемость студентов: влияние школы / Е. А. Попова, М. В. Шеина. Текст: непосредственный // Современный университет между глобальными вызовами и локальными задачами: сборник материалов 7-й Международной конференции Российской ассоциации исследователей высшего образования / под ред. Д. В. Козлова, Н. Г. Малошенок. Москва: Изд. дом Высш. шк. экономики, 2016. С. 183–187.

4. Татусь, К. Ю. Влияние родительской семьи на успеваемость студентов / К. Ю. Татусь, С. В. Кузьмина. Текст: непосредственный // Молодой ученый 2016. № 9.4 (113.4). С. 69–72.

5. Богданов, Е. П. О прогнозировании успеваемости студентов по результатам ЕГЭ и атрибутам социального статуса / Е. П. Богданов, А. В. Суханов. Текст: непосредственный // Актуальные направления научных исследований XXI века: теория и практика. 2015. Т. 3, № 7–3(18–3). С. 382–386.

6. Русаков, С. В. Прогнозирование успеваемости студентов первого курса с помощью дерева решений на основе их результатов сдачи ЕГЭ / С. В. Русаков, Н. Н. Накарякова. Текст: непосредственный // Новые информационные технологии в образовании: материалы Международной научно-практической конференции, 26 февр.–2 марта 2018 г., Екатеринбург / Рос. гос. проф.-пед. ун-т. Екатеринбург, 2018. С. 589–594.

7. Русаков, С. В. Нейросетевая модель прогнозирования группы риска по успеваемости студентов первого курса / С. В. Русаков, О. Л. Русакова, К. А. Посохина. Текст: непосредственный // Современные информационные технологии и ИТ-образование. 2018. Т. 14, № 4. С. 799–807.