

УДК 330.4:311.14

М. В. Боченина

ФГБОУ ВО «Санкт-Петербургский государственный экономический университет»,
Санкт-Петербург, e-mail: m-bochenina@yandex.ru

Ю. В. Нерадовская

ФГБОУ ВО «Санкт-Петербургский государственный экономический университет»,
Санкт-Петербург, e-mail: neradovskaya.yu@unecon.ru

Л. Б. Курмашева

Северо-Казахстанский университет им. М. Козыбаева, Петропавловск,
e-mail: lb_kurmasheva@mail.ru

В. П. Куликова

Северо-Казахстанский университет им. М. Козыбаева, Петропавловск,
e-mail: v4lentina@mail.ru

УСПЕШНОСТЬ ОБУЧЕНИЯ В ВУЗЕ: СТАТИСТИКО-ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ

Ключевые слова: логит-модель, множественная регрессия, нейронная сеть, успешность обучения, уровень бедности.

В статье представлены результаты исследования, проведенного на основе данных, собранных в региональном университете Республики Казахстан, целью которого являлось проведение статистического анализа выявления факторов, оказывающих заметное влияние на успешность окончания университета студентами. Под успешностью в работе понималась вероятность не быть отчисленным до окончания срока обучения. В результате исследования были выделены три группы факторов, отражающие индивидуальные характеристики абитуриентов, студентов и характеристики уровня социально-экономического развития региона, в котором расположен вуз. Однако, как и ожидалось, наиболее существенным оказалось отношение студентов к обучению, а именно их успеваемость. В качестве инструментария для построения статистико-эконометрических моделей процесса успешного окончания университета были использованы методы дискретного выбора, регрессионного анализа временных рядов и нейронных сетей.

M. V. Bochenina

State University of Economics, St. Petersburg, e-mail: m-bochenina@yandex.ru

Yu. V. Neradovskaya

State University of Economics, St. Petersburg, e-mail: neradovskaya.yu@unecon.ru

L. B. Kurmasheva

Manash Kozybayev North Kazakhstan University, Petropavlovsk,
e-mail: lb_kurmasheva@mail.ru

V. P. Kulikova

Manash Kozybayev North Kazakhstan University, Petropavlovsk, e-mail: v4lentina@mail.ru

SUCCESS IN HIGHER EDUCATION: STATISTICAL AND ECONOMETRIC ANALYSIS

Keywords: logit model, multiple regression, neural network, learning success, poverty level.

The article presents the results of a study conducted based on data collected in a regional university of the Republic of Kazakhstan, the purpose of which was to conduct a statistical analysis to identify factors that have a significant impact on the success of students to graduate. Success in work was understood as the probability of not being expelled before the end of the training period. As a result of the study, three groups of factors were identified, reflecting the individual characteristics of applicants, students, and characteristics of the level of socio-economic development of the region in which the university is located. However, as expected, the most significant factor was the students' attitude to learning, namely their academic performance. The methods of discrete choice, regression analysis of time series and neural networks were used as tools to build statistical-econometric models of the process of successful graduation.

Введение

Вопросы управления университетами, в том числе отдельными протекающими в них процессами, в последнее время приобретают все большую популярность. Связано это, прежде всего, с возрастающей конкуренцией между вузами, с желанием занимать верхние позиции в мировых рейтингах. Алгоритмы управления отдельными процессами в разных вузах в корне отличаются между собой.

Согласно А. Лукичеву, директору центра компетенций НТИ по нейротехнологиям, «вузам поставили задачу коммерциализировать их внутренние разработки, и они стали наполовину бизнес-структурами. Каждый вуз – это отдельная планета, и все работают по-разному: у кого-то отдельно свои лаборатории, у кого-то своя внутренняя цифровая система...» [6].

Руководство любой организации при принятии управленческих решений, как правило, опирается на достоверные и актуальные данные. Это касается и объектов образования. Поэтому разработка новых методов, моделей управления процессами в организациях высшего образования в настоящее время интересуются многие ученые. А сейчас, в условиях пандемии COVID-19 и межвузовской конкуренции проблема набора контингента и увеличения количества выпускников особенно остро встает перед руководством образовательных учреждений. Рынок труда непрерывно меняется, профессии, которые были востребованы в последние десятилетия, становятся ненужными, появляется потребность в специалистах новых направлений. В настоящее время многие университеты инвестируют в прогнозную аналитику, которая обеспечивается данными, генерируемыми учебной деятельностью в Интернете, а также опирается на доступ к внешним консультационным экспертам.

Получение высшего образования – один из важнейших факторов повышения качества человеческого капитала как для любого государства, так и для отдельно взятых регионов. Вопросы, связанные с набором контингента и выпуском бакалавров и специалистов, являются одними из самых важных как для вузов, так и для рынка труда рассматриваемого региона. Государство выделяемыми грантами старается отрегулировать «провалы» образовательного рынка. Растущая конкуренция на рынке об-

разовательных услуг, отсутствие надежных прогнозов и не совершенность механизмов по определению данного показателя, делает исследования по данной тематике особенно актуальными.

Рассмотрев публикации по теме использования методов и приложений прогнозирования, становится ясно, что интерес к данной области возрос в 10 раз за последние 40 лет [9].

Цель исследования провести анализ факторов, влияющих на вероятность обучающего успешно окончить обучение в университете.

Материалы и методы исследования

В качестве рабочей гипотезы предполагалось существование зависимости между социально-экономическими показателями региона и успешным окончанием обучения по программам бакалавриата региональных вузов. Аналогичное исследование проведено В.Ф. Пугач, в результате которого было выявлено повышение устойчивости студенческого контингента с увеличением конкурса при поступлении в вуз [5]. В качестве инструмента исследования был использован коэффициент корреляции рангов Ч. Спирмена.

В работе [8] в результате исследования было установлено, что академическая успешность студентов взаимосвязана с добросовестностью и ответственностью. В работе [3] определены статистически значимые детерминанты успешной успеваемости студентов.

При проведении исследования в качестве исходных данных были использованы статистические данные по региональному вузу Республики Казахстан (НАО Северо-Казахстанский университет имени М. Козыбаева) и материалы Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан, опубликованные на его официальном сайте [2]. Это показатели, отражающие индивидуальные характеристики абитуриентов, индивидуальные характеристики студентов при обучении в вузе и характеристики уровня социально-экономического развития Северо-Казахстанской области, в которой расположен исследуемый университет.

Первая группа показателей включает средний балл аттестата и место постоянного жительства, которое представляет собой бинарную переменную, равную единице при проживании по месту нахождения учебного

заведения, и ноль при проживании в другой местности.

Вторая группа – накопленный в процессе обучения рейтинг, который в Казахстане называется GPA (Grade Point Average) и характеристика формы обучения: с оплатой (коммерческая форма) или на бесплатной основе, которая в Казахстане определяется грантом, а в России является бюджетной формой обучения. Последний показатель является бинарной переменной, принимающей значение единица для бюджетной формы и ноль для коммерческой.

Третья группа показателей – среднедушевой номинальный денежный доход населения, доля населения, имеющего доходы ниже величины прожиточного минимума и уровень безработицы.

В качестве результирующей рассматривалась бинарная переменная, принимающая значение единица, если студент был отчислен до окончания обучения.

Данные представлены за период 2006–2017 гг., объем выборки составил 14877 студентов.

Характер данных предполагает использование в качестве инструмента анализа моделей бинарного выбора.

Так как данные содержат независимые выборки за 12 периодов времени, по каждому периоду были рассчитаны средние значения количественных показателей, а бинарные переменные преобразованы в количественные в виде относительных показателей доли. В частности, результирующая переменная приняла форму доли отчисленных до окончания обучения, объясняющие переменные «место постоянного жительства» – доли проживающих по месту нахождения учебного заведения, «характеристика формы обучения» – доли обучающихся на бюджете.

Проведенные преобразования данных позволили применить для анализа успешности обучения модель регрессии по временным рядам.

В качестве альтернативного метода анализа больших данных был использован метод Левенберга-Маркара для построения нейронной сети, согласно которому обучение происходит «по эпохам» [1]. Данный метод предполагает, что ошибка построенной сети определяется за всю эпоху обучения, а параметры модели будут изменяться, когда сеть пройдет все элементы обучающей части выборки [7]. С этой целью, имеющуюся выборку разделили на три части:

обучающую, контрольную и тестовую. Для обучения, то есть построения самой модели, использовали 70% данных, для текущей оценки состояния обучения – 15% данных, оставшиеся 15% данных были использованы для процесса тестирования качества построенной модели.

В настоящее время нейронные сети используются для решения большого класса задач, в том числе прогнозирования. Обученная сеть способна обобщать информацию и давать хорошие результаты даже на данных, которые не использовались в процессе обучения [4].

Для обработки данных были использованы инструменты *Ms Excel*, *Gretl*, *Matlab R2017b*.

Результаты исследования и их обсуждение

Проведенный предварительный корреляционный анализ межфакторных связей показал наличие мультиколлинеарности, что привело к сокращению количества объясняющих переменных. В частности, индивидуальные характеристики абитуриента имеют статистически значимую связь с характеристиками обучающегося. Студенты бюджетной формы обучения имеют в среднем более высокие баллы GPA. Тесно связаны между собой показатели третьей группы, отражающие уровень социально-экономического развития региона.

При построении моделей бинарного выбора наилучшей оказалась логит-модель следующего вида:

$$P(\text{otch} = 1) = \frac{1}{1 + e^{-(5,19 - 0,06bedn - 2,36GPA + \varepsilon)}}$$

где $P(\text{otch} = 1)$ – вероятность быть отчисленным до окончания обучения, $bedn$ – доля населения, имеющего доходы ниже величины прожиточного минимума, ε – случайный остаток модели.

Все параметры модели статистически значимы, уравнение значимо (критерий отношения правдоподобия: Хи-квадрат (2) = 4488,09 [0,0000]), индекс *McFadden R²* равен 0,39, предсказательная сила модели составляет 92,8%.

Таким образом, с ростом доли бедных вероятность быть отчисленным уменьшается. Предельный эффект для среднего значения данного фактора составил (-0,003). То есть с изменением доли бедных на один

процентный пункт от среднего значения вероятность отчисления студента до окончания обучения изменяется в противоположную сторону на 0,003 при неизменном среднем балле *GPA*.

Очевидным образом вероятность отчисления повышается при уменьшении баллов текущей аттестации. Предельный эффект равен (-0,144), что показывает величину роста вероятности быть отчисленным при снижении балла *GPA* на единицу от среднего при неизменном уровне бедности.

Модель с этими же объясняющими переменными, построенная по временным рядам, имеет следующий вид:

$$dotch = 0,58 - 0,003 \cdot bedn - 0,137 \cdot \overline{GPA} + \varepsilon$$

$$(t) \quad (8,7) \quad (-6,2) \quad (-6,7),$$

где *dotch* – доля студентов, отчисленных в учебном году, \overline{GPA} – средний балл для по выборке студентов для рассматриваемого года.

В скобках указаны фактические значения критерия Стьюдента, критическое значение составило 2,3 на уровне значимости 0,05. Скорректированный коэффициент

детерминации равен 0,80. Уравнение регрессии значимо ($p - value = 2,7 \cdot 10^{-8}$). Анализ случайных остатков отсутствие автокорреляции (критерий Дарбина-Уотсона $d_w = 2,11$ при критических значениях 0,81 и 1,58) и нормальность их распределения ($p - value = 0,97$).

Полученные коэффициенты практически полностью совпадают с предельными эффектами, рассчитанными по логит-модели, однако логит-модель имеет более высокую предсказательную силу.

Таким образом, обе модели показали, что основным фактором, который определяет вероятность не попасть в число отчисленных, является накопленный в процессе обучения рейтинг (*GPA*).

Следующим этапом исследования успешности обучения в университете было построение нейронной сети с использованием обучающей выборки по всем показателям трех описанных выше групп. Результаты обучения определялись среднеквадратической ошибкой (*MSE*) и значением коэффициента корреляции (*R*) между выходами сети и целевыми значениями (таблица).

Значения результатов обучения нейронной сети

	Объем выборки	Среднеквадратическая ошибка	Коэффициент корреляции
Обучающая выборка	10413	0,061	0,692
Контрольная выборка	2232	0,051	0,728
Тестовая выборка	2232	0,065	0,638

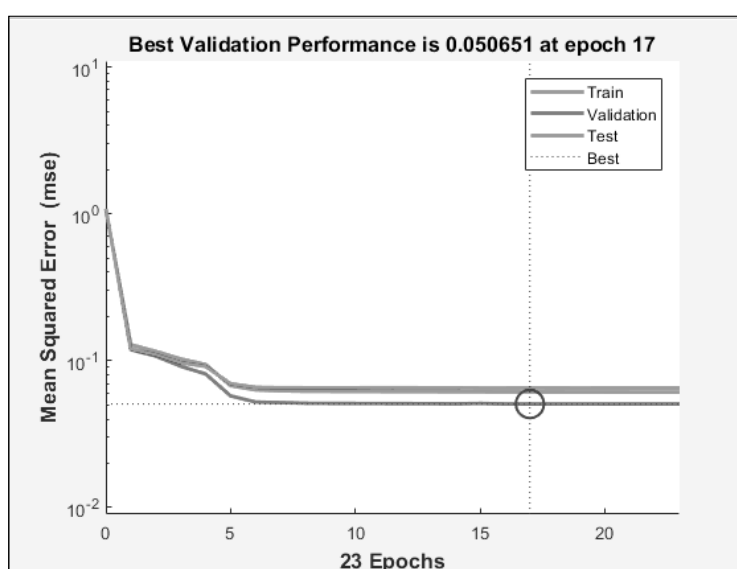


График эффективности обучения

Сравнение трех построенных моделей показало, что несмотря на то, что в нейронной сети были использованы все предложенные в начале исследования показатели, результат, полученный с её использованием, оказался хуже, чем у моделей регрессии. Так, коэффициент детерминации на разных выборках составлял от 0,41 до 0,53. Предсказательная сила нейронной сети, согласно тестовой выборке, составила 90,0%, что несколько ниже, чем по модели дискретного выбора (92,8%).

В целом можно заметить уменьшение ошибки во время процесса обучения на каждой эпохе (рисунок). Обучение было остановлено, когда ошибка на проверочном множестве перестала уменьшаться.

Заключение

Анализ факторов, оказывающих влияние на успешное окончание обучения в универ-

ситете, проведенный по данным регионального университета в Республике Казахстан, показал, что наиболее существенным можно назвать отношение студентов к обучению, их прилежание и, соответственно, полученные оценки по дисциплинам. Социально-экономические показатели, в том числе низкий уровень благосостояния населения, являются одним из стимулов, мотивирующих студентов к ответственному отношению к учебному процессу, что подтверждено результатами эконометрического моделирования. Молодые люди стремятся после окончания обучения перейти в более высокие, лучше материально обеспеченные слои общества. Таким образом, рабочая гипотеза была подтверждена результатами эконометрического моделирования, как и очевидный факт, что успешность окончания университета напрямую зависит от отношения студента к учебе.

Библиографический список

1. Балацкий Е.В., Юревич М.А. Использование нейронных сетей для прогнозирования инфляции: новые возможности // Вестник УрФУ. Серия экономика и управление. 2018. Т. 17. № 5. С. 823–838. DOI: 10.15826/vestnik.2018.17.5.037.
2. Динамика основных социально-экономических показателей. Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан. [Электронный ресурс]. Режим доступа: <https://stat.gov.kz/region/264023/dynamic>.
3. Канапухин П.А., Коротких В.В., Щекунских С.С. Статистический анализ факторов академической успеваемости (на примере обучающихся экономического факультета ФГБОУВО «ВГУ») // Вестник Воронежского государственного университета. Серия: Экономика и управление. 2020. № 2. С. 27–44. DOI: 10.17308/econ.2020.2/2899.
4. Осовский С. Нейронные сети для обработки информации / пер. с польского И.Д. Рудинского. М.: Финансы и статистика, 2002. 344 с.
5. Пугач В.Ф. Конкурс при поступлении в вузы, прием и выпуск: выявление взаимосвязи // Высшее образование в России. 2014. № 3
6. Пандемия цифрового образования. [Электронный ресурс]. Режим доступа: <https://vc.ru/education/153144-pandemiya-cifrovogo-obrazovaniya>.
7. Ферцев А.А. Ускорение обучения нейронной сети для распознавания изображений с помощью технологии NVIDIA CUDA // Вестник Самарского государственного технического университета. Сер. Физ.-мат. науки. 2012. Вып. 1(26). С. 183–191. DOI: 10.14498/vsgtu990.
8. Шапоров А.М., Исаева Е.Р., Тюсова О.В., Ванчакова Н.П., Кулик В.В. Анализ факторов, влияющих на успешность обучения студентов медицинского вуза // Ученые записки университета Лесгафта. 2019. № 6 (172).
9. Maximilian Zellner, Ali E. Abbas, David V. Budescu and Aram Galstyan. A survey of human judgement and quantitative forecasting methods. DOI: 10.1098/rsos.201187.