

ПРИМЕНЕНИЕ АНАЛИТИКИ БОЛЬШИХ ДАННЫХ В ОНЛАЙН-ОБУЧЕНИИ ИНОСТРАННЫХ СТУДЕНТОВ, ИЗУЧАЮЩИХ РУССКИЙ ЯЗЫК В ЦЕЛЯХ ИХ АДАПТАЦИИ

Погодин И.М., аспирант, Российский институт стандартизации

Садыхбеков А.Д., аспирант, ФГБУ «Московский городской педагогический университет»

В статье представлено исследование, цель которого заключается в изучении эффективности применения аналитики больших данных для мониторинга успеваемости и социально-экономической адаптации иностранных студентов, изучающих русский язык в онлайн-формате.

Методы исследования включают использование регрессионного анализа на основе данных 130 студентов для выявления ключевых факторов, влияющих на результаты обучения и социально-экономической адаптации.

Исследование определило, что наиболее значимыми факторами являются частота входа в систему и результаты онлайн-тестирования, в то время как продолжительность онлайн-обучения и субъективные оценки преподавателей оказались менее существенными. Разработанная модель продемонстрировала 85%-ную точность в прогнозировании успешности адаптации студентов. Результаты исследования подтверждают эффективность использования аналитики больших данных в онлайн-образовании и предлагают практические рекомендации по улучшению качества обучения русскому языку и адаптации иностранных студентов.

Ключевые слова: большие данные, онлайн-обучение русскому языку, иностранные студенты, социально-экономическая адаптация, регрессионный анализ, прогнозирование успеваемости, образовательная аналитика, мониторинг успеваемости, интеграция иностранных студентов.

ВВЕДЕНИЕ

С анализ современных исследований показывает, что использование аналитики больших данных в образовании активно развивается. Согласно исследованиям аналитика данных позволяет выявлять индивидуальные траектории обучения [1, 2], прогнозировать успеваемость [3, 4] и улучшать процессы адаптации иностранных студентов [5, 6]. Эти методы особенно актуальны в условиях онлайн-обучения, где важна обратная связь в реальном времени [7, 8].

Цель данного исследования заключается в оценке эффективности применения аналитики больших данных для мониторинга успеваемости и социально-экономической адаптации иностранных студентов, изучающих русский язык в онлайн-формате.

Большие данные, развивающиеся на фоне суперкомпьютеров и облачных вычислений, открывают новые возможности для обработки огромных объемов информации в сфере высшего образования. В контексте обучения иностранных студентов русскому языку и их социально-экономической адаптации аналитика больших данных предоставляет уни-

кальные инструменты для оптимизации образовательного процесса и улучшения результатов обучения.[1]

Для иностранных студентов, изучающих русский язык, обучающая аналитика может не только раскрыть механизмы процесса обучения, но и помочь в их социально-экономической адаптации. Анализ данных позволяет рекомендовать индивидуальные траектории обучения, учитывающие не только языковые навыки, но и культурные особенности, что способствует более эффективной интеграции студентов в новую среду. Для преподавателей и администраторов эта технология предоставляет возможность оценивать эффективность курсов и образовательных программ, а также разрабатывать стратегии поддержки иностранных студентов в их адаптации к новой социально-экономической реальности [2, 3].

АНАЛИЗ МОДЕЛИ

В данном исследовании используется регрессионный анализ для изучения факторов, влияющих на результаты онлайн-обучения русскому языку и социально-экономическую адаптацию иностранных студентов. Предполагается,

что зависимая переменная Y отражает не только языковые навыки, но и уровень адаптации студента, а независимые переменные X_1, X_2, \dots представляют различные аспекты обучения и адаптации (1). [4]

Применяемая модель представлена в виде линейного регрессионного уравнения:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (1)$$

где Y – уровень адаптации, X_n – аспекты обучения и адаптации, β_n – коэффициенты регрессии, ε – случайная ошибка.

Выбор переменных

В процессе обучения иностранных студентов русскому языку и их адаптации учебное поведение и социальная

активность студента играют ключевую роль. Основываясь на этом, были выбраны следующие переменные, представленные в табл. 1.

Результаты и анализ

Результаты вычислений показаны в табл. 2 после импорта данных учебной группы (30 студентов) в регрессионную модель.

Интерпретация

Среди независимых переменных значения X_3 и X_4 показали наибольшую значимость для успешной социально-экономической адаптации студентов. Коэффициент X_4 оказался наиболее влиятельным, что подчеркивает важность регулярного тестирования не только языковых навыков, но и понимания культурного контекста. Онлайн-тесты стали эффективным инструментом оценки как лингвистических

Таблица 1

Описание переменных

ПЕРЕМЕННАЯ	ДИАПАЗОН ЗНАЧЕНИЙ	ОПИСАНИЕ	ИСТОЧНИК ДАННЫХ
X1	60–100	Представляет способность к самостоятельному обучению, выраженную через предварительный просмотр материалов курса и повторение материала на учебной платформе	Формируется на основе оценок за чтение и перевод текста, выставленных преподавателями
X2	1–30 000	Представляет общее время, затраченное на просмотр небольших видео и участие в сессиях вопросов и ответов	Автоматически регистрируется на онлайн-платформе обучения
X3	1–450	Количество входов в систему в рамках курса	Полученные выходные данные с онлайн-платформы обучения
X4	0–100	Сумма баллов, полученных в различных онлайн-тестах	Рассчитывается автоматически на основе результатов тестов, выполненных на платформе

Таблица 2

Первая модель регрессии

ПЕРЕМЕННАЯ	КОЭФФИЦИЕНТ	СТ. ОШИБКА	t-СТАТИСТИКА	ВЕРОЯТНОСТЬ (p-value)
С (Константа)	7.009081	23.029459	0.256501	0.1467
X1	0.200529	0.258738	0.560173	0.4434
X2	-0.000229	0.036872	-0.855364	0.3198
X3	0.047482	0.029939	1.868429	0.0646
X4	0.559167	0.140276	4.936536	0.0017
R-КВАДРАТ	0.66862	СРЕДНЕЕ ЗНАЧЕНИЕ ЗАВИСИМОЙ ПЕРЕМЕННОЙ	66.13793	
Скорректированное R-квадрат	0.622894	Стандартное отклонение зависимой переменной	10.01686	
Ст. ошибка регрессии	6.175118	Критерий Акаике	6.615014	
Сумма квадратов остатков	896.0169	Критерий Шварца	6.843619	
Логарифм правдоподобия	-90.8952	Критерий Ханнана-Куинна	6.671502	
F-статистика	12.92441	Статистика Дурбина-Уотсона	2.202163	
p-значение (F-статистика)	0.000001			

компетенций, так и уровня адаптации студентов к новой среде. Переменная X3, отражающая частоту входа в систему, помогла преподавателям отслеживать не только учебную активность, но и общую вовлеченность студентов в процесс адаптации. Это позволило своевременно выявлять студентов, нуждающихся в дополнительной поддержке. Показатель посещаемости также продемонстрировал сильную корреляцию с успешностью социально-экономической интеграции.

При этом влияние переменных X1 и X2 оказалось менее значительным. Особенно это касается X2, где близкий к

нулю коэффициент показал, что простое время пребывания в системе не гарантирует успешной адаптации. Переменная X1, основанная на субъективных оценках преподавателей, также не продемонстрировала существенного влияния на общий процесс адаптации студентов.[4]

Модификация и эмпирический анализ

После модификации модели были исключены переменные X1 и X2. Результаты вычислений представлены в табл. 3.

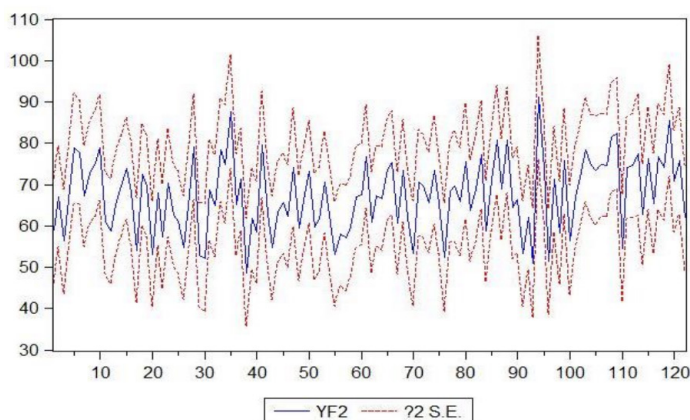
Таблица 3 показывает, что после исключения переменных X1 и X2 коэффициент переменной X3 увеличился до 0.040013

Таблица 3

Первая модель регрессии

ПЕРЕМЕННАЯ	КОЭФФИЦИЕНТ	СТ. ОШИБКА	t-СТАТИСТИКА	ВЕРОЯТНОСТЬ (p-value)
С (Константа)	19.67912	6.51775	3.020003	0.0056
X3	0.040013	0.016998	2.301481	0.0296
X4	0.602119	0.094514	6.297303	0.0000

R-КВАДРАТ	0.66862	СРЕДНЕЕ ЗНАЧЕНИЕ ЗАВИСИМОЙ ПЕРЕМЕННОЙ	66.13793
Скорректированное R-квадрат	0.643129	Стандартное отклонение зависимой переменной	10.01686
Ст. ошибка регрессии	5.984029	Критерий Акаике	6.514033
Сумма квадратов остатков	930.9958	Критерий Шварца	6.655177
Логарифм правдоподобия	-91.4501	Критерий Ханнана-Куинна	6.560031
F-статистика	26.23005	Статистика Дурбина-Уотсона	2.199258
p-значение (F-статистика)	0.000001		



Прогноз второй регрессионной модели
 Ось X: Фактические показатели адаптации
 Ось Y: Предсказанные показатели адаптации

с уровнем значимости $p = 0.0296$, что подтверждает усиление влияния данной переменной на успешность социально-экономической адаптации студентов. Эти изменения объясняются уменьшением мультиколлинеарности между X1, X2 и X3. Важно отметить, что данные представлены для выборки из 130 студентов, а ключевые параметры модели представлены в табл. 3. Прогнозируемые значения адаптации (YF) представлены на рисунке с интервалами, рассчитанными по формуле $\square \pm 2 \times \square \cdot \square$.

В табл. 4 представлены показатели, которые оценивают качество предсказания модели:

- **среднеквадратичная ошибка (Root Mean Squared Error – RMSE):** это мера того, насколько сильно предсказания модели отклоняются от фактических значений. Чем меньше это значение, тем точнее модель. В данном случае ошибка около 8.94 указывает на то, что предсказанные оценки в среднем отклоняются от реальных оценок на 8.94 балла;
- **средняя абсолютная ошибка (Mean Absolute Error – MAE):** это среднее значение абсолютных отклонений

между предсказанными и реальными значениями. MAE, равная 6.77, означает, что в среднем ошибка модели составляет около 6.77 балла. Этот показатель дает представление о том, насколько модель «ошибается» в своих предсказаниях;

- **средняя абсолютная процентная ошибка (Mean Abs. Percent Error – MAPE):** это процентная ошибка, которая показывает средний процент отклонений между предсказанными и фактическими значениями. В данном случае MAPE равна 11.25%, что означает, что в среднем модель ошибается на 11.25% в предсказаниях итоговых оценок;
- **коэффициент неравенства Тейла (Theil Inequality Coefficient):** этот коэффициент оценивает качество предсказания, принимая во внимание как общую ошибку, так и систему смещения в предсказаниях. Чем ниже этот коэффициент, тем лучше модель. Значение 0.066612 указывает на хорошее качество предсказания.
- **пропорция смещения (Bias Proportion):** пропорция смещения показывает, какую долю ошибки можно объяснить систематическим отклонением в предсказаниях (например, предсказания всегда слишком высоки или слишком низки). Значение 0.035983 указывает на очень небольшое смещение модели, что хорошо, поскольку это означает, что модель не склонна к систематическим ошибкам;
- **пропорция дисперсии (Variance Proportion):** пропорция дисперсии показывает, какая доля ошибки связана с различиями между предсказаниями. Низкое значение 0.012142 указывает на то, что модель не слишком чувствительна к вариациям в данных;
- **пропорция ковариации (Covariance Proportion):** это значение показывает, сколько из ошибки модели объясняется ковариацией между предсказаниями и реальными значениями. Высокая пропорция ковариации (0.952019) означает, что большая часть ошибки зависит от взаимосвязи между предсказаниями и фактическими значениями, а не от случайных факторов;
- **коэффициент Тейла U2 (Theil U2 Coefficient):** это дополнительная мера точности модели, основанная на анализе неравенства Тейла. Коэффициент, близкий к 0, указывает на хорошее предсказание, и значение 0.572505 указывает на средний уровень точности модели;
- **симметричная MAPE (Symmetric MAPE):** симметричная версия MAPE исправляет проблему, когда ошибки на больших значениях имеют больший вес. Это также показывает среднюю процентную ошибку, но с учетом симметричности, и составляет 10.55%, что подтверждает довольно хорошую точность модели.

YF2 на рисунке показывает предсказанные значения успешности социально-экономической адаптации студентов. Доверительные интервалы для этих прогнозов рассчиты-

Таблица 4

Результаты качества показания модели YF2

ПОКАЗАТЕЛЬ	ЗНАЧЕНИЕ
Прогноз	YF2
Фактические данные	Y
Диапазон прогноза	1–120
Включенные наблюдения	130
Среднеквадратичная ошибка	8.940001
Средняя абсолютная ошибка	6.773011
Средняя абсолютная процентная ошибка	11.25003
Коэффициент неравенства Тейла	0.066612
Пропорция смещения	0.035983
Пропорция дисперсии	0.012142
Пропорция ковариации	0.952019
Коэффициент Тейла U2	0.572505
Симметричная MAPE	10.55028

ваются на основе стандартной ошибки регрессии (S.E. of Regression).

Формула для расчета верхнего и нижнего пределов интервала:

$$\text{Upper / Lower} = YF \pm 2 \times S.E. \quad (2)$$

где значение S.E. составляет 5.984029. Анализ показал, что фактические показатели адаптации 110 из 130 студентов (85%) попали в предсказанные интервалы, что свидетельствует о высокой точности модели в прогнозировании успешности социально-экономической адаптации иностранных студентов.

Эти результаты основаны на данных таблицы 3, где указаны средние значения и интервалы прогнозируемых значений. Высокая точность модели подтверждается низкими значениями ошибки прогноза (MAE, RMSE), представленными в таблице 4.

ЗАКЛЮЧЕНИЕ

Большие данные играют ключевую роль в преподавании русского языка иностранным студентам и их социально-экономической адаптации. Они позволяют не только контролировать эффективность обучения, но и отслеживать процесс адаптации студентов к новой среде.

Анализ показал, что регулярность участия в онлайн-курсах, успешность в тестах и активное участие в культурных мероприятиях являются ключевыми факторами успешной

адаптации и освоения языка. Это позволяет преподавателям и администраторам разрабатывать более эффективные стратегии поддержки иностранных студентов.

Однако остаются проблемы, связанные со сбором и анализом данных. Необходимо развивать цифровую грамотность преподавателей и их способность анализировать данные, а также налаживать сотрудничество с IT-специалистами для создания эффективных систем анализа данных.

Данное исследование подтверждает значимость аналитики больших данных для онлайн-обучения русскому языку и социально-экономической адаптации иностранных студентов. Использование регрессионного анализа позволяет эффективно прогнозировать не только академические результаты, но и успешность адаптации студентов. Это открывает новые перспективы для эмпирических исследований в области преподавания иностранных языков и поддержки иностранных студентов в их интеграции в новую социально-экономическую среду.

Список использованных источников и литературы

1. Меркулова Т.А. Адаптация иностранных студентов к обучению в вузе культуры и искусств // Вестник Московского государственного университета культуры и искусств. – 2010. – № 4. – С. 184–186.
2. Янь Жуйтин. Модель смешанного обучения китайских студентов русскому языку с использованием платформы WeChat (уровень В1) // Диссертация. – 2023.
3. Aijuan, C., Qian, Z. Research on China's studying abroad situation and countermeasures // 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering. – Bali, Indonesia, 2011. – Pp. 75–78. – DOI: 10.1109/URKE.2011.6007843.
4. Шилова З.В. Теория вероятностей и математическая статистика: учебное пособие / З.В. Шилова, О.И. Шилов. – Киров: Изд-во ВГГУ, 2015. – 158 с.
5. Галямова Э.Х., Матвеев С.Н., Киселев Б.В. Оценка эффективности цифрового симулятора педагогической деятельности средствами статистического анализа // Высшее образование сегодня. – 2024. – № 4. – С. 45–54.
6. Косова Е.А. Курс по организации доступного образовательного веб-контента: анализ результатов онлайн-обучения // Открытое образование. – 2021. – Т. 25, № 2. – С. 29–40. DOI: 10.21686/1818-4243-2021-2-29-40.
7. Акаике Х. Новый взгляд на идентификацию статистической модели // IEEE Transactions on Automatic Control. – 1974. – Т. 19, № 6. – С. 716–723.
8. Шварц Г. Оценка размеров модели // Annals of Statistics. – 1978. – Т. 6, № 2. – С. 461–464.
9. Хэйр Дж.Ф., Блэк У.К., Бабин Б. Дж., Андерсон Р.Е. Многомерный статистический анализ. – Пирсон, 2010.
10. Хосмер Д.В., Лемешоу С. Прикладной логистический регрессионный анализ. – Уайли, 2000.
11. Табачник Б.Г., Фиделл Л.С. Использование многомерной статистики. – Пирсон, 2013.

APPLICATION OF BIG DATA ANALYTICS IN ONLINE LEARNING FOR FOREIGN STUDENTS STUDYING RUSSIAN LANGUAGE FOR ADAPTATION PURPOSES

Pogodin I.M., PhD student of the Russian Standardization Institute

Sadikhbekov A.D., PhD student of the Moscow City Pedagogical University

The article presents a study aimed at examining the effectiveness of using big data analytics to monitor academic performance and socio-economic adaptation of foreign students studying Russian in an online format.

The research methods include the use of regression analysis based on data from 130 students to identify key factors influencing learning outcomes and socio-economic adaptation.

The study revealed that the most significant factors are the frequency of system logins and online testing results, while the duration of online learning and teachers' subjective assessments proved to be less significant. The developed model demonstrated 85% accuracy in predicting students' adaptation success.

The research results confirm the effectiveness of big data analytics in online education and provide practical recommendations for improving the quality of Russian language teaching and the adaptation of foreign students.

Keywords: big data, online Russian language learning, foreign students, socio-economic adaptation, regression analysis, performance prediction, educational analytics, academic performance monitoring, integration of foreign students.

References

1. Merkulova, T.A. Adaptation of foreign students to studying at a university of culture and arts // Bulletin of Moscow State University of Culture and Arts. – 2010. – No. 4. – Pp. 184–186.
2. Yan, Ruiting. A blended learning model for teaching Russian to Chinese students using the WeChat platform (level B1): Dissertation for the degree of Candidate of Pedagogical Sciences. – 2023.
3. Aijuan, C., Qian, Z. Research on China's studying abroad situation and countermeasures // 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering. – Bali, Indonesia, 2011. – P. 75–78. DOI: 10.1109/URKE.2011.6007843.
4. Shilova, Z.V. Theory of probability and mathematical statistics: a textbook / Z.V. Shilova, O.I. Shilov. – Kirov: Vyatka State University Publishing House, 2015. – 158 p.
5. Galyamova, E.Kh., Matveev, S.N., Kiselev, B.V. Assessment of the effectiveness of a digital simulator of pedagogical activity using statistical analysis methods // Higher Education Today. – 2024. – No. 4. – Pp. 45–54.
6. Kosova, E.A. A course on creating accessible educational web content: an analysis of online learning results // Open Education. – 2021. – Vol. 25, No. 2. – Pp. 29–40. DOI: 10.21686/1818-4243-2021-2-29-40.
7. Akaike, H. A new look at the statistical model identification // IEEE Transactions on Automatic Control. – 1974. – Vol. 19, No. 6. – Pp. 716–723.
8. Schwarz, G. Estimating the dimension of a model // Annals of Statistics. – 1978. – Vol. 6, No. 2. – Pp. 461–464.
9. Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. Multivariate data analysis. – Pearson, 2010.
10. Hosmer, D.W., Lemeshow, S. Applied Logistic Regression. – Wiley, 2000.
11. Tabachnick, B.G., Fidell, L.S. Using Multivariate Statistics. – Pearson, 2013.