

УДК 004.65:37.062

РАЗРАБОТКА АНАЛИТИЧЕСКОГО ПОДХОДА НА ОСНОВЕ БОЛЬШИХ ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ФАКТОРОВ УСПЕХА СТУДЕНТОВ



Т.В. Казак

*Заведующий кафедрой инженерной психологии и эргономики, доктор психологических наук Республики Беларусь, доктор психологических наук Российской Федерации, профессор, БГУИР
kazak@bsuir.by*



А.В. Свороб

*Студентка БГУИР, кафедра инженерной психологии и эргономики
nastyasvorob@gmail.com*



А.Н. Василькова

*Старший преподаватель кафедры инженерной психологии и эргономики, БГУИР
a.vasilkova@bsuir.by*

Т.В. Казак

Заведующий кафедрой инженерной психологии и эргономики, доктор психологических наук Республики Беларусь, доктор психологических наук Российской Федерации, член-корреспондент Международной академии психологических наук, профессор.

Образование: Высшее; Аспирантура, специальность: 19.00.03 - психология труда, инженерная психология, эргономика; Докторантура, специальность: 19.00.03 - психология труда, инженерная психология, эргономика.

Область профессиональных интересов / исследований: Психология труда. Инженерная психология. Эргономика. Психология управления. Юридическая психология. Социальная и организационная психология. Клиническая психология.

А.В. Свороб

Студентка кафедры инженерной психологии и эргономики БГУИР.

Область профессиональных интересов / исследований: языки программирования, искусственный интеллект, технологии виртуальной реальности.

А.Н. Василькова

Старший преподаватель кафедры инженерной психологии и эргономики.

Образование: 2007 - МГВРК по специальности «Программное обеспечение информационных технологий»,

2022 - магистратура БГУИР по специальности «Охрана труда и эргономика».

Область профессиональных интересов / исследований: языки программирования, искусственный интеллект, технологии виртуальной реальности.

Аннотация. Снижение отсева студентов в высшем образовании играет важную роль в выполнении основной миссии и финансовом благополучии учебного заведения. Доступность источника больших данных из системы управления обучением (LMS) может быть проанализирована, чтобы помочь в решении проблемы отсева. Целью данного исследования является использование интегрированной методологии научного исследования дизайна (DSR) для разработки и оценки нового аналитического решения на основе больших

данных (*BDAS*) в качестве артефакта поддержки принятия решений в сфере образования. *BDAS* как артефакт *DSR* использует подходы искусственного интеллекта для прогнозирования потенциальных студентов, находящихся в группе риска. Выявление студентов, находящихся в группе риска, помогает своевременно вмешаться в учебный процесс, чтобы улучшить успеваемость студентов и повысить процент их удержания.

Ключевые слова: Искусственный интеллект, исследование на основе дизайна, система поддержки принятия решений, большие данные, информационная система, машинное обучение, образование.

Введение. В последнее время искусственный интеллект широко и эффективно применяется в компьютерной сфере. В литературе отмечаются преимущества и возможности применения искусственного интеллекта (ИИ) в сфере образования. Несколько примеров применения: аналитика данных, прогнозирование зачисления студентов, система рекомендаций для карьерного роста или управления ресурсами, адаптивное репетиторство, прогнозирование готовности студентов к трудоустройству, мониторинг и прогнозирование успеваемости студентов или выявление проблемных студентов. Существующие исследования не фокусируются на больших данных *LMS* для прогнозирования академической успеваемости на ранних этапах обучения. В большинстве исследований использовались данные, полученные в ходе транзитного обучения или полностью онлайн, и лишь в небольшом количестве исследований изучались данные, полученные при взаимодействии студентов с *LMS* в смешанном обучении. Кроме того, в большинстве существующих исследований не подчеркивается важность выявления студентов группы риска на ранних этапах обучения. Необходимо разработать аналитическое решение, работающее в режиме реального времени, для раннего выявления студентов, подверженных риску неуспеваемости, в среде смешанного обучения, чтобы своевременно предложить им стратегии и меры по исправлению ситуации для поддержания академического прогресса студентов. Кроме того, большинство исследований, связанных с методологией исследования и созданием и оценкой артефактов *DSR*, являются недостаточными, учитывая, что эти исследования не используют интегрированную методологию *DSR* и *DBR* для планирования исследования по проектированию и разработке артефакта; эти исследования используют подходы анализа больших данных, но не используют *DSR* или *DBR* или интегрированную парадигму *DSR*; эти исследования не оценивают артефакты *DSR* в соответствии с их сложностью. Тем не менее, существующая литература может быть использована для экстраполяции для достижения цели данного исследования, таким образом, формируя основу данного исследования.

Основная часть. На начальных этапах нашей комплексной методологии исследования *DSR* был проведен обширный системно-аналитический обзор литературы и мета-анализ (*SLRM*) по применению технологий ИИ в высшей школе в отношении академической успеваемости студентов. Систематический обзор литературы направлен на понимание тенденций применения технологий на основе ИИ в широком спектре, связанных с мониторингом и прогнозированием академической успеваемости студентов, а также на выявление различных алгоритмов ИИ и процесса разработки моделей ИИ. *SLRM* был проведен с использованием системы *PRISMA* с определением поискового протокола, включающего критерии включения и исключения, и предоставлением богатых результатов. В *SLRM* были выделены этапы, алгоритмы и метрики оценки, использованные в исследованиях. Эти алгоритмы и метрики оценки легли в основу проектирования и разработки *BDAS*.

Целью проектирования и разработки *BDAS* является обучение и оценка прогностической модели на основе классифицированных данных для прогнозирования академической успеваемости студентов. Прогностическая модель должна быть достаточно точной, чтобы выявлять студентов, которым грозит неуспеваемость. Прогнозирование может помочь преподавателям в реализации стратегий, направленных

на повышение эффективности обучения студентов и улучшение их академической успеваемости. *BDAS* может быть интегрирована в учебные курсы для своевременного и точного определения академического прогресса студентов, особенно студентов группы риска. Своевременное выявление студентов, находящихся в группе риска, способствует более раннему вмешательству для улучшения их академической успеваемости. Общая вычислительная модель состоит из сбора данных, предварительной обработки данных, анализа данных с помощью алгоритмов и оценки. Эта типовая модель адаптируется к каждой итерации этапа проектирования и разработки *BDAS*. В каждой итерации использовались различные методы предварительной обработки и различные алгоритмы для достижения цели *BDAS*.

В случае с большими данными в сфере образования *LMS* генерирует большое количество данных в режиме реального времени. Прогностическая модель *BDAS* обучается на наборе исторических данных о взаимодействии студентов с *LMS*, как показано в данном исследовании. Для сбора поступающих больших данных и создания сегментов данных используется распределенная платформа обработки больших данных, например *Apache Kafka* и *Spark*.

Эти небольшие сегменты *Fahd and Miah Journal of Big Data* Распределенная платформа обработки больших данных используется для сбора входящих больших данных и создания сегментов данных. Эти партии больших данных классифицируются *BDAS* для выявления студентов, находящихся в группе риска, а *ML*-модели принимают все входные данные одновременно для создания выходных данных, что невозможно в *BDAS* из-за большого объема и высокой производительности. скорость больших данных. Существуют различные подходы к решению этой проблемы и применению алгоритмов ИИ для разработки модели на больших образовательных данных, такие как технологии параллельной обработки, высокопроизводительная вычислительная инфраструктура, платформы обработки данных для разделения данных.

В данном исследовании предлагается использовать платформу обработки данных для архитектуры метода *BDA*. Однако, в данном исследовании основное внимание уделяется проектированию, разработке и оценке *BDAS*, а не архитектурной среде *BDAS*.

Артефакт *DSR*, основанный на ИИ, является сложным артефактом и разрабатывается в соответствии с требованиями и целями, определенными на предыдущих этапах. Подходы к проектированию, разработанные на основе контекстуальных знаний и общей практики, приводят к улучшению дизайна артефакта. В данном исследовании использовались два набора итераций для проектирования и разработки *BDAS* в качестве прогностической модели на основе существующих в литературе подходов: Прогностическая модель на основе *ML*; Прогностическая модель на основе *DL*. На этом этапе мы применяем алгоритмы *ML* и *DL* для проектирования и разработки прогностических моделей на основе *ML* и *DL* в качестве артефактов *DSR* для точного определения потенциальных студентов с риском неуспеваемости из набора данных, основанных на взаимодействии студентов с *LMS*. Итерационный подход в этой фазе обеспечивает непрерывное улучшение построения *DSR*-артефакта путем оценки различных показателей эффективности с помощью матрицы путаницы в каждой итерации. Эти показатели эффективности различных алгоритмов ИИ в каждой итерации сравниваются для выбора лучшей прогностической модели. *BDAS* как артефакт *DSR* строится из серии задач, состоящих из сбора данных, предварительной обработки данных, анализа данных с помощью алгоритмов искусственного интеллекта, оценки и принятия успешного решения. Все эти задачи предназначены для разработки и оценки прогностических моделей на основе *ML* и *DL*. Рабочий процесс обучения артефакта на основе ИИ показан на рисунке 1.

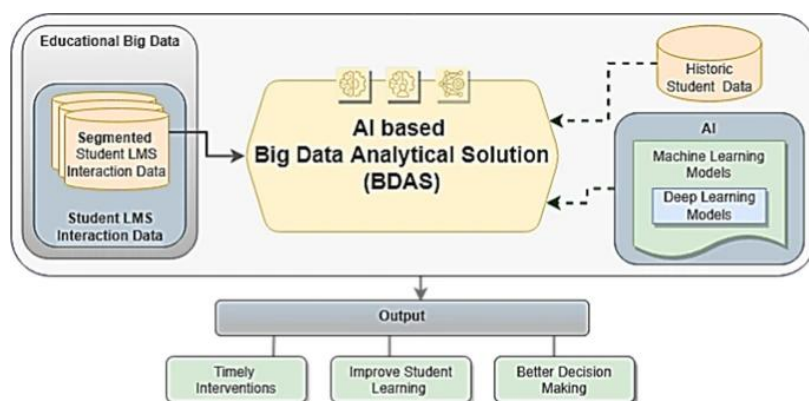


Рисунок 1. Рабочий процесс обучения артефакта на основе ИИ

Для обучения прогностической модели в данном исследовании был использован свободно распространяемый набор данных из репозитория *UCI ML*, включающий 230 318 примеров действий и взаимодействий студентов с *LMS*.

На первой итерации для обучения и оценки прогностической модели используются пять древовидных алгоритмов *ML* с супервизией (*J48*, *Random Forest*, *OneR*, *Decision Stump*, *NBTree*). Эти древовидные алгоритмы используют серию решений «если – то» для создания высокоточных, легко интерпретируемых прогнозов, чтобы выявить потенциальных студентов, подверженных риску неуспеваемости. Для дальнейшей настройки к преобразованному набору данных применяется метод ансамбля бустеров. Прогностическая модель обучается и тестируется с помощью *k*-кратной перекрестной валидации на обучающих и тестовых данных с использованием пяти вышеупомянутых алгоритмов *ML* с супервизией итеративно. На последнем этапе сравниваются показатели эффективности всех прогностических моделей, основанных на пяти алгоритмах *ML*, чтобы выбрать наиболее точную прогностическую модель для построения *BDAS*. При реализации *BDAS* в реальном времени для получения и сегментации потока больших данных из *LMS* в реальном времени будет использован фреймворк обработки данных, например *Apache spark*, который разделит большие данные на небольшие партии для обработки и классификации предиктивной моделью *BDAS*.

Во второй итерации непрерывного совершенствования артефакта, основанного на искусственном интеллекте, используются два различных метода предварительной обработки данных для изменения распределения классов и дополнения набора данных, чтобы устранить последствия дисбаланса данных. Алгоритмы *DL* состоят из нейронных сетей с несколькими слоями дифференцируемых нелинейных узлов. Три *DL*-алгоритма – *Long Short-term Memory (LSTM)*, *Multi-layer Perceptron (MLP)* и *Sequential Model (SM)* – были применены для обучения дополненного набора данных, что показало более высокую точность классификации модели предсказания и уменьшило количество ложных предсказаний. Более высокая точность классификации и уменьшение количества ложных предсказаний означают низкую вероятность того, что не будут выявлены учащиеся, не входящие в группу риска, что соответствует цели общего описания *BDAS* как артефакта *DSR*.

Заключение. В исследовании представлена интеграция двух исследовательских методологий *DSR* и *DBR* на основе ключевых сходств между ними для проектирования, создания и оценки нового артефакта *DSR* под названием *BDAS*. Методологический взгляд формирует соответствующую исследовательскую парадигму для проектирования, разработки и оценки артефакта *BDAS*, который может быть внедрен для повышения академической успеваемости с помощью стратегий своевременного вмешательства для

тех, кто находится под угрозой неуспеваемости, и поддержки принятия более эффективных решений.

Список литературы

- [1] Алджохани О. Всесторонний обзор основных исследований и теоретических моделей удержания студентов в высшем образовании. М: High Educ Stud. 2016. – 200 с.
[2] Бир К., Лоусон К. Проблема отсева студентов в высшем образовании: альтернативный взгляд. М.: J Furth High Educ. 2017. – 773 с.
[3] Отоо-Артур Д., ван Зил Т. Масштабируемая гетерогенная структура больших данных для систем электронного обучения. М.: IEEE; 2020. – 448 с.
[4] Анг Л-М, Ге Ф, Сенг К. Большие образовательные данные и аналитика: обзор, архитектура и проблемы. М.: IEEE Access. 2020. – 414 с.

Авторский вклад

Казак Тамара Владимировна – руководство и постановка задачи исследования аналитического подхода на основе больших данных для прогнозирования факторов успеха студентов.

Василькова Анастасия Николаевна – постановка задачи исследования, описание принципа работы Big Data в улучшении эффективности прогнозирования факторов успеха студентов, анализ полученных результатов, формирование структуры статьи.

Свороб Анастасия Вячеславовна – тестирование программного средства, описание принципов использования прогнозирования факторов успеха студентов, формирование структуры статьи.

DEVELOPMENT OF ANALYTICAL BIG DATA APPROACH FOR PREDICTIONS OF SUCCESS FACTORS STUDENTS

T.V. Kazak

Head of the Department of Engineering Psychology and Ergonomics, doctor of psychological sciences of the Republic of Belarus, doctor of psychological sciences of the Russian Federation, Professor, Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus.

A.V. Svorob

BSUIR student, Department of Engineering Psychology and Ergonomics

A.N. Vasilkova

Senior Lecturer, Department of Engineering Psychology and Ergonomics, BSUIR

Abstract. Reducing student attrition in higher education plays an important role in achieving the core mission and financial health of the institution. The availability of big data source from the learning management system (LMS) can be analyzed to help address the attrition problem. The purpose of this study is to use an integrated design scientific research (DSR) methodology to develop and evaluate a new big data analytics solution (BDAS) as an educational decision support artifact. BDAS as an artifact DSR uses artificial intelligence approaches to predict potential at-risk students. Identifying students who are at risk helps to intervene early in the educational process to improve student performance and increase retention rates.

Keywords: Artificial intelligence, design-based research, decision support system, big data, information system, machine learning, education.