

43. DATA SCIENCE ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

*Столбун Е. А., студентка гр. 272301, Полоско Е.И., аспирант,
Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Ефремов А.А. – канд. экон. наук, доцент каф. ЭИ

Аннотация. Данная статья посвящена способам применения Data Science для улучшения производительности обучения студентов. Названы причины, по которым наука о данных полезна для сферы образования. Предложены идеи для проектов, предназначенных для эффективной учебы. Рассмотрены возможные проблемы, связанные с цифровизацией обучения и приведены способы их решения.

Ключевые слова. Data Science, цифровизация, образование, машинное обучение.

В условиях цифровизации всех областей жизни значительно изменился подход к обучению. Современные информационные технологии значительно повлияли на доступность образования и

методах управления учебными процессами. Цифровизация упрощает доступ к учебным ресурсам, повышению уровня интерактивности учебы и распространению онлайн-образования.

Одним из наиболее эффективных методов повышения эффективности обучения является Data Science. Data Science — это междисциплинарная область на стыке статистики, математики и машинного обучения, которая охватывает все этапы работы с данными. Она предполагает исследование и анализ больших массивов информации и ориентирована на получение практических результатов [1].

В современном образовании существует ряд способов применения Data Science. Например, одной из актуальных проблем является персонализация обучения. Студенты имеют разные уровни знаний, способности и предпочтения. Именно Data Science позволяет адаптировать учебные материалы и методики под индивидуальные потребности каждого учащегося. Благодаря анализу учебной активности студента, его успеваемости, скорости усвоения материала можно составить уникальный учебный план. Также при помощи машинного обучения можно разработать систему адаптивной системы оценки знаний, которая корректирует сложность вопросов на основе знаний студентов.

Кроме того, Data Science может помочь в выявлении неуспевающих студентов. При помощи методов машинного обучения можно прогнозировать отметку студента за экзамен и сдачу зачета на основе таких факторов, как текущие отметки, посещаемость и др. Это позволяет вовремя выявлять студентов с трудностями в учебе для оказания необходимой им помощи. Data Science может применяться и для того, чтобы вычислить вероятность выпуска студентов из университета и нахождения ими трудоустройства.

Также наука о данных может применяться для составления расписания. Для этого необходимо выявить учебные предпочтения студентов, такие как учеба в первую/вторую смену, количество занятий в день и т.д. Затем к полученным данным следует применить алгоритмы кластеризации. Алгоритмы кластерного анализа позволяют определить группы (кластеры) данных более схожих друг с другом, чем с остальными данными и выявить ранее незамеченные закономерности [2]. К подобным алгоритмам относятся K-means clustering, Expectation-maximization algorithm, DBSCAN и многие другие. Ниже представлено краткое описание работы K-means clustering, наиболее популярного метода кластеризации.

Работа алгоритма начинается с выбора K центроид. Центроида — точка, являющаяся центром кластера. Каждой точке из набора данных присваивается ближайший центроид. Затем координаты центроидов пересчитываются. Данные операции прекратятся тогда, когда координаты центроидов перестанут меняться, точки остаются в тех же кластерах или достигнуто максимальное число итераций [3].

Кластерный анализ позволяет разделить студентов на группы в зависимости от их стиля обучения и сформировать расписание занятий для каждой группы. Также он может быть полезен для разбиения студентов на группы для совместного выполнения учебных проектов.

Более того, Data Science может применяться для оптимизации учебных программ путем прогнозирования потребностей рынка труда, предварительно собрав данные о вакансиях.

Для анализа наиболее востребованных навыков можно использовать методы Natural Language Processing. К примеру, при помощи Named Entity Recognition можно идентифицировать языки программирования, инструменты и технологии, упомянутые в описаниях вакансий. Named Entity Recognition автоматически идентифицирует именованные объекты в тексте и классифицирует их по заранее определенным категориям [4]. Такой инструмент, как анализ тональности текста способен определить, какие навыки упоминаются с позитивной эмоциональной окраской [5]. Это поможет лучше понять, какие компетенции важно развивать.

Для прогнозирования востребованности определенных навыков в будущем можно применить такие модели машинного обучения, как дерево решений, случайный лес, метод опорных векторов и др.

В соответствии с полученной информацией можно внести изменения в учебные планы, чтобы сделать их более актуальными. Это поможет учебным заведениям идти в ногу со временем и развивать в студентах необходимые рынку труда компетенции.

На сегодняшний день множество образовательных сервисов уже применяют Data Science. Например, Coursera анализирует поведение студентов и предлагает персонализированные рекомендации по курсам и материалам. Khan Academy создает учебные планы для каждого ученика на основе их уровня знаний и прогресса. Duolingo адаптирует упражнения по изучению иностранных языков в зависимости от успехов и ошибок пользователя. Вышеописанные сервисы применяют машинное обучение для достижения данных целей.

Несмотря на то, что внедрение информационных технологий в образование имеет множество преимуществ, есть и определенные проблемы.

Одной из ключевых проблем является недостаточная конфиденциальность данных студентов в условиях цифрового обучения. В условиях сбора, хранения и обработки больших объемов информации

возникает угроза утечки или неправомерного использования конфиденциальных данных. Развитие методов шифрования и технологий обеспечения кибербезопасности становится важной задачей для образовательных учреждений.

Другой проблемой, связанной с цифровизацией образования, является цифровое неравенство, которое проявляется в неравном доступе студентов к высокоскоростному интернету и современным технологиям. Для решения этой проблемы можно использовать анализ данных о доступности цифровых ресурсов и разработать стратегии по устранению цифрового неравенства.

Кроме того, качество образовательных материалов в условиях цифровизации может оказаться недостаточным из-за разнообразия онлайн-курсов и учебных ресурсов. Data Science может быть применена для анализа обратной связи от студентов, оценки эффективности учебных материалов и создания рекомендательных систем для выбора наиболее подходящих курсов.

Стоит отметить, что недостатком применения Data Science в образовании является возможность искажения результатов анализа данных из-за неправильной интерпретации или недостаточной квалификации специалистов. В случае неправильного анализа данных, принятие ошибочных решений в образовательном процессе может привести к негативным последствиям для студентов и учебных заведений. Поэтому необходимо обеспечить высокий уровень компетентности специалистов, занимающихся анализом данных в образовании.

Также необходимо учитывать возможность возникновения этических проблем. Например, использование алгоритмов машинного обучения для прогнозирования успеваемости студентов может привести к созданию предвзятых моделей, основанных на расовой, социальной или другой дискриминации. Поэтому важно разрабатывать этические стандарты и строго следить за их соблюдением при использовании Data Science в образовании.

Более того, при применении Data Science в образовании возможна потеря человеческого фактора в образовательном процессе. Алгоритмы и модели машинного обучения могут помочь оптимизировать процессы и улучшить результаты, однако они не могут заменить важное личное взаимодействие между учителями и учениками. Эмоциональная поддержка, мотивация, индивидуальный подход к каждому студенту — все это сложно передать через алгоритмы. Поэтому необходимо находить баланс между использованием технологий и сохранением гуманитарного компонента в образовании.

Таким образом, Data Science является мощным методом для повышения эффективности учебного процесса. Ее внедрение в образование способствует адаптации учебного процесса под потребности каждого учащегося и созданию более гибких и эффективных образовательных систем. Несмотря на выявленные недостатки цифровизации образования, использование Data Science для обучения открывает новые возможности для студентов и преподавателей.

Список использованных источников:

1. *Data Science for analyzing and improving educational processes* / Shadi Aljawarneh, Juan A. Lara // *Journal of Computing in Higher Education*, 2021. – P. 545-550.
2. *Data clustering: application and trends* / Gbeminiyi John Oyewole, George Alex Thopil // *Artificial Intelligence Review*, 2023. – P. 6439-6475.
3. *Выбор алгоритмов кластеризации* / А.С. Чочиева, И.И. Пилецкий // *Шестая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня»*, Минск, Республика Беларусь, 20-21 мая 2020 года. – С.281-293.
4. *Named Entity Recognition in English Text* / Rachna Jain [et.al.] // *Journal of Physics: Conference Series* 1712 012013, 2020. – P.1-6.
5. *Анализ тональности текста методами машинного обучения* / Самигулин Т.Р., Джурабаев А.Э.У. // *Научный результат. Информационные технологии*. – Т.6, №1, 2021. – С. 55-62.