

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
КРАСНОЯРСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ
УНИВЕРСИТЕТ им. В.П. Астафьева

(КГПУ им. В.П. Астафьева)

Институт /факультет Институт Математики, физики и информатики
Кафедра Информатики и информационных технологий в образовании

Изимбетов Нурлан Орныкбаевич

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**Разработка цифрового сервиса для анализа качества обучения на основе
технологий Big data**

Направление подготовки: 44.04.01 Педагогическое образование

Направленность (профиль) образовательной программы: Технологии
цифровизации образовательной деятельности (с применением сетевой формы) с
Казахским национальным педагогическим университетом им. Абая

ДОПУСКАЮ К ЗАЩИТЕ

Заведующий кафедрой

д-р. пед. наук, проф.

Пак Н.И.

Руководитель магистерской программы

д-р. пед наук, проф., зав. каф. ИиИТО

Пак Н.И.

Научный руководитель канд. пед. наук,

д-р. пед наук, проф.

Пак Н.И.

Обучающийся

Изимбетов Н.О.

Дата защиты «09» июня 2022 г.

Оценка (прописью) _____



Красноярск, 2022

Реферат

Актуальность исследования: В век информационных технологий, особенно с бурным развитием цифровых технологий, у каждого пользователя Интернета стало накапливаться значительное количество информации, что в свою очередь привело к появлению технологии Big Data. И многие научные исследования показали, что использование этой технологии в реальном образовании повышает качество учебного процесса. Используя возможности этой новой технологии, можно создать цифровой сервис предиктивного анализа качества образования. Поэтому данная тема является одним из важнейших вопросов при разработке цифрового сервиса предиктивного анализа качества образования.

Проблема исследования заключается в вопросе - каким образом можно использовать технологии Big Data для повышения качества и эффективности учебно-педагогического процесса?

Цель исследования: Разработка модели обучающегося, обеспечивающей построение индивидуальных образовательных траекторий в соответствии с индивидуальными качествами и способностями обучающегося, расчет прогноза поведения обучающегося в процессе обучения, учет его индивидуальных характеристик и образовательных результатов с применением технологии Big Data.

Задачи исследования:

- Разработка цифрового сервиса предиктивного анализа качества образования обучающихся на основе научно-методических исследований;
- Выявление ситуаций, связанных с большими данными, при разработке цифровых сервисов для предиктивного анализа качества образования;
- Особенности анализа и применения больших данных при разработке цифровых сервисов для предиктивного анализа качества образования;

- Выбор технологий больших данных при разработке цифровых сервисов для предиктивного анализа качества образования;

- Разработка модели цифрового сервиса предиктивного анализа качества образования;

- Внедрение цифрового компьютерного сервиса прогнозного анализа качества образования.

Теоретическая значимость работы заключается в обосновании возможности интеграции и систематизации информации, собранной по данной теме. **Практическая значимость** работы заключается в создании цифрового сервиса предиктивного анализа качества образования на основе технологий Big Data, показывающего возможность эффективного использования современной технологии Big Data при разработке цифровых сервисов для предиктивного анализа качества образования.

Основные результаты исследования: Анализ возможностей новых информационных технологий в использовании технологий больших данных в отечественных и зарубежных исследованиях. Применение математического моделирования, технологий программирования при разработке цифровых сервисов предиктивного анализа качества образования. На основе технологий Big Data разработана модель цифрового сервиса прогнозного анализа качества образования в университете, включающая в себя структуру, компоненты, функции и результаты. Разработка цифрового сервиса предиктивного анализа качества обучения учащихся на основе технологии Big Data.

Оглавление

Введение	5
1. Опыт применения прогнозного анализа качества образования в университете.....	7
1.1. Использование Big Data и цифровой аналитики в университетском образовании.....	7
1.2. Отбор Big Data технологий прогнозного анализа качества образования.....	10
1.3. Выявление ситуаций, связанных с большими данными, при создании цифрового сервиса прогнозного анализа качества образования.....	17
2. Модели прогнозного анализа качества образования в университете на основе технологий Big Data.....	25
2.1. Алгоритмы оценки участия студентов в учебном процессе.....	25
2.2. Модели прогнозирования учебной "траектории" студентов.....	35
2.3. Критерии прогнозного анализа качества образования в университете.....	42
3. Методические основы создания и использования цифрового сервиса прогнозного анализа качества образования.....	47
3.1. Особенности анализа и использования больших данных при создании цифрового сервиса прогнозного анализа качества образования.....	47
3.2. Архитектура цифрового сервиса прогнозного анализа качества образования в университете.....	52
3.3. Реализация цифрового сервиса прогнозного анализа качества образования в университете на компьютере.....	56
3.4. Методические указания по использованию цифрового сервиса прогнозного анализа качества образования.....	63
Заключение.....	70
Библиографический список.....	72

Введение

Актуальность исследования: В век информационных технологий, особенно с бурным развитием цифровых технологий, у каждого пользователя Интернета стало накапливаться значительное количество информации, что в свою очередь привело к появлению технологии Big Data. И многие научные исследования показали, что использование этой технологии в реальном образовании повышает качество преподавания. Используя возможности этой новой технологии, можно создать цифровой сервис предиктивного анализа качества образования. Поэтому данная тема является одним из важнейших вопросов при разработке цифрового сервиса предиктивного анализа качества образования.

Проблема исследования заключается в вопросе - каким образом можно использовать технологии Big Data для повышения качества и эффективности учебно-педагогического процесса?

Цель работы: Разработка модели обучающегося, обеспечивающей построение индивидуальных образовательных траекторий в соответствии с индивидуальными качествами и способностями обучающегося, расчет прогноза поведения обучающегося в процессе обучения, учет его индивидуальных характеристик и образовательных результатов с применением технологии Big Data.

Задачи исследования:

- Разработка цифрового сервиса предиктивного анализа качества образования обучающихся на основе научно-методических исследований;
- Выявление ситуаций, связанных с большими данными, при разработке цифровых сервисов для предиктивного анализа качества образования;
- Особенности анализа и применения больших данных при разработке цифровых сервисов для предиктивного анализа качества образования;

- Выбор технологий больших данных при разработке цифровых сервисов для предиктивного анализа качества образования;
- Модель цифрового сервиса предиктивного анализа качества образования;
- Внедрение цифрового компьютерного сервиса прогнозного анализа качества образования.

Теоретическая значимость работы заключается в обосновании возможности интеграции и систематизации информации, собранной по данной теме. **Практическая значимость** работы заключается в создании цифрового сервиса предиктивного анализа качества образования на основе технологий Big Data, показывающего возможность эффективного использования современной технологии Big Data при разработке цифровых сервисов для предиктивного анализа качества образования.

1. ОПЫТ ПРИМЕНЕНИЯ ПРОГНОЗНОГО АНАЛИЗА КАЧЕСТВА ОБРАЗОВАНИЯ В УНИВЕРСИТЕТЕ

1.1. Использование Big Data и цифровой аналитики университетском образовании

Началом активного использования термина «Большие данные» стал специальный выпуск журнала Nature за 2008 год под названием «Как научные технологии, способные работать с большими объемами данных, могут повлиять на будущее». В нем представлены материалы о быстром росте информации [5].

В настоящее время Большие данные описывают набор из семи символов («семь V»), которые важны для понимания сложности этих задач обработки и анализа данных: Объем, Скорость, Разнообразие, Изменчивость, Достоверность, Ценность, Визуализация. Рассмотрим эти признаки подробнее:

1. Объем более 150 Гб в сутки.

2. Информация о скорости обновляется с высокой скоростью и требует интеллектуальных технологий для ее обработки в режиме реального времени.

3. Разнообразие. Разнообразие данных, которые могут быть структурированными, неструктурированными или частично структурированными. Каждый из этих типов данных требует различного анализа и соответствующих инструментов.

4. Изменчивость. Изменчивость потоков данных, которая может привести к пикам и падениям, что приводит к трудностям управления и требует мощных технологий обработки.

5. Достоверность набора данных и результаты его анализа.

6. Ценность (ценность или значимость) определяется эффективностью использования результатов их анализа для принятия решений.

7. Визуализация является частью анализа больших данных и делает их доступными для человеческого восприятия.

Существует четыре метода анализа больших данных: описательная аналитика, диагностическая аналитика, прогнозная аналитика и предписывающая аналитика. Это шаги по обеспечению зрелости аналитики, каждый из которых сокращает дистанцию между этапами анализа и откликом на данные. Описательная аналитика анализирует то, что произошло. Диагностическая аналитика анализирует не только события, но и их причины. Прогнозная аналитика предсказывает возможные результаты на основе выявленных тенденций и статистических моделей. Предписывающая аналитика позволяет находить оптимальные решения на основе прогнозной аналитики и часто основана на аналитике машинного обучения и нейронных сетях. Этот тип аналитики требует прочной основы, основанной на трех других типах аналитики.

Большие данные стали использоваться для решения задач в области образовательной аналитики [6-7].

Источник больших данных в университетах:

1. Персональные данные;
2. Информация о взаимодействии студентов с системами электронного обучения (электронные учебники, онлайн-курсы и т. д.);
3. Информация об эффективности учебных материалов (с каким контентом взаимодействует обучающийся, результаты взаимодействия, образовательные результаты и т.д.);
4. Административные (общесистемные) данные (посещаемость, отсутствие по определенным причинам и т.п.).

Рассмотрим примеры использования больших данных и цифровой аналитики в зарубежном и российском вузовском образовании.

В 2013 г. в Университете Ноттингем-Трент, Англия, была внедрена интерактивная система описательного анализа результатов студентов в виде дашборда, на котором отображаются данные об участии студентов в процессе обучения. Панель предназначена для снижения отсева студентов, улучшения посещаемости занятий и повышения чувства принадлежности к университетскому сообществу. Панель, доступная для студентов, преподавателей и кураторов (тьюторов), показывает участие каждого студента в сравнении со своими сверстниками: частоту использования библиотеки, информацию о пройденных курсах, посещаемость и другие показатели обучения. Таким образом, любой студент может увидеть свою деятельность и сравнить себя со своими однокурсниками, чтобы понять, насколько он вовлечен в учебный процесс и университетскую жизнь в целом, а также на что следует обратить больше внимания. Если студент не проявляет признаков активности в течение двух недель, платформа отправит уведомления кураторам, чтобы они могли быстро связаться и поддержать студента. Через три года после внедрения системы результаты университетского опроса показали, что 72% первокурсников

использовали этот студенческий дашборд Big Data, что вдохновило их на увеличение времени учебы [8].

Американский университет Пердью запустил систему предиктивного анализа, которая оценивает риск отсева каждого студента. Значение этого показателя периодически отправляется координатору курса и студенту. В университете отмечают, что использование интерактивных больших данных позволило улучшить результаты обучения и снизить риск отсева студентов [8].

Также интересен опыт Университета Карнеги-Меллона, еще одного вуза в США, который создал онлайн-платформу с обучающими курсами, предоставляющими студентам подробные отчеты о своей успеваемости. Результаты его работы показали, что возможность просмотра подробных обзоров и анализа своего прогресса и учебной деятельности почти удвоила время прохождения онлайн-курсов [8].

Российские вузы уже анализируют свои большие данные об образовательном процессе. Для этого они не только адаптируют имеющиеся на рынке коммерческие инструменты, но и разрабатывают собственные технические решения. Команда исследователей из Московского государственного педагогического университета разработала систему, которая с точностью 71% предсказывает успеваемость студента в следующем семестре. В модели используются самые разные данные: экзаменационные баллы, результаты предыдущих сессий, информация об участии студента в общественной деятельности и получении книг из университетской библиотеки и электронных библиотечных систем. МГПУ планирует дальнейшее развитие своей системы, в том числе интеграцию с рекомендуемой цифровой платформой. Если прогноз успеваемости плохой или хороший, его может увидеть только сотрудник вуза. В дальнейшем это будет напрямую связано с индивидуальными занятиями учащихся, которые смогут оценивать свои достижения и создавать собственную траекторию решения задач [9].

Уральский государственный педагогический университет разработал модель прогнозирования посещаемости. В качестве отправной точки он использует информацию об индивидуальных особенностях учащихся: темперамент, готовность к работе, пересечение сферы интересов студента с предметом, мотивацию, способность к самоорганизации и другое. Дальнейшее обучение в нейронной сети позволяет разработать рекомендации по развитию студентов и составить расписание, которое поможет повысить посещаемость [10].

1.2. Отбор Big Data технологий прогнозного анализа качества образования.

Одним из важнейших аспектов развития системы образования является способность делать обоснованные выводы о необходимости изменения подходов или предпринимаемых действий. В системе образования непрерывно генерируется и накапливается значительный объем данных, и вопрос систематической работы с этими данными широким кругом субъектов образования можно назвать одним из важнейших на сегодняшний день. Большие данные могут быть мощным инструментом для преобразования обучения, переосмысления подходов, сокращения давних пробелов и адаптации методов для повышения эффективности системы образования. Задача описания технологии работы с большими данными, направленной на развитие систем образования путем выявления закономерностей в системе образования, остается весьма актуальной.

Исследования в отечественной и зарубежной литературе по проблеме использования технологии Big Data в системе образования достаточно разбросаны.

Сегодня большие данные становятся языком общения образовательных организаций, стремящихся к совершенствованию технологий принятия стратегических и тактических решений [11].

Стоит обратить внимание на содержательный анализ тем более трех тысяч научных статей в области «компьютеры и образование», проведенный О.Заваки-Рихтером и С.Латчемом. Анализ за последние 40 лет информационные причины можно сгруппировать по четырем хронологическим периодам: развитие и рост компьютерного обучения (1976-1986 годы); мультимедийное обучение (1987-1996 годы); сетевые технологии для организации совместного обучения (1997-2006 годы); онлайн обучение (2007-2016 годы). Таким образом, можно отметить, что сегодня научное сообщество уделяет большое внимание проблемам онлайн-обучения, способам повышения его эффективности, которые невозможно найти без всестороннего анализа Big Data, собираемого по результатам обучения [12].

С. Виейра, П. Парсонс, В. Берд характеризуют результаты анализа литературы по поиску данных в образовании [13]. Результаты исследования показывают, что визуальное обучение на уровне одного класса группы мало внимания уделяется средствам анализа. Редким параметром анализа данных является демография и предыдущие учебные достижения. Н. Буньямин, У. Б. Мет, П. М. Аршад проанализировали наиболее часто используемые в области интеллектуального анализа образовательных данных методы классификации для прогнозирования академических достижений учащихся [14].

Другим аспектом исследования больших данных является инфраструктура собранных данных. Таким образом, Ф. А. Де Альмейда Нето и А. Кастро разработали модель, согласно которой данные, созданные в результате взаимодействия между пользователями с учетом онлайн-платформ, на которых проводятся образовательные мероприятия, и сама платформа выбирается, собирается и хранится в локальных базах данных [15]. Затем локальные базы данных собираются и группируются в глобальные базы данных.

Существуют сравнительные исследования средств прогнозирования учебных достижений. Таким образом, Д. Буэнаньо-Фернандес и С. Луан-Мора анализируют три инструмента с открытым исходным кодом (RapidMiner, Knime

и Weka), используемых в интеллектуальном образовании [16]. Х. Ю и С. Ву говорят о целевых пользователях и группах участников, участвующих в создании больших данных в образовании, а также анализируют цели использования собранных для них данных [17].

М. Фарохмер, С.О. Фатиме описывают требования к Big Data по формуле 3V: большой объем, скорость обработки и разнообразие параметров [18].

Отдельными аспектами данной области исследования являются проблемы, связанные с достижением образовательных результатов. EDM (Educational Data Mining) описывается как инструмент повышения эффективности электронного обучения. Таким образом, М. Насири, Б. Минаи, Ф. Вафай рассматривают модель прогнозирования академического дохода путем наблюдения и поддержки первокурсников [19]. В работах О. Москосо-Цеа, М. Фицкайно, С. Луйан-Мора методы Big Data оцениваются по двум показателям эффективности образования: отчисление студентов и выпуск студентов [20].

Для структурирования процессов управления Big Data в образовании можно выделить пять взаимосвязанных групп процессов (рис. 1.1) [21]:

1. целеполагание: определение цели и задач исследования;
2. Планирование: выбор источников информации, процедур получения данных, алгоритмов обработки информации;
3. сбор данных: организация сбора данных в единую базу данных;
4. анализ показателей: анализ полученных данных, определение способов представления результатов;
5. корректировка: разработка практических мер регулирования;
6. Завершение: утверждение законности.



Рис.1.1. Схема процессов управления Big Data в образовании

Обобщая группы и источники процессов управления по направлениям Big Data, можно сформулировать ряд свойств накопленной базы данных, позволяющих повысить эффективность использования Big Data в образовании.

1. Частичная независимость

Каждый сегмент базы данных имеет локальное управление данными. В то же время каждый сегмент является составной частью всей базы данных, которую нельзя рассматривать как отдельную небольшую базу данных с отдельным набором процедур и правил [21].

2. Непрерывность

Возможность получения данных из любого сегмента базы данных, даже если данные в этом сегменте используются для других процессов.

3. Прозрачность доступа

При наличии прав на доступ к данным аналитик не должен учитывать параметры конкретного места размещения информации. Доставка данных осуществляется автоматически встроенными средствами.

4. Умножение

Умножение данных-это процесс передачи данных из одной базы данных в другую. Возможность умножения должна позволять данным различных систем интегрироваться друг с другом.

5. Распределенные запросы

Собранные данные должны быть получены через распределенные запросы, то есть запросы, параллельные нескольким сегментам базы данных.

6. Бесплатные инструменты

Любые программно-аппаратные решения могут быть средством обработки данных.

Таким образом, Big Data в образовании как технология характеризуется своеобразием, структурированием процессов управления, системой направлений сбора данных, а также свойствами собираемой базы данных, позволяющей повысить эффективность использования Big Data в образовании [21].

Для хранения больших данных требуется большой объем дисковой памяти. Для решения данной проблемы используется кластерная технология NAS (Network Area Storage), предусматривающая подключение накопителей информации непосредственно к локальной или распределенной компьютерной сети с использованием протокола TCP/IP. Такая сеть позволяет пользователям хранить файлы на серверах NAS и обмениваться ими через браузер или его сетевой адрес. Кластерная инфраструктура NAS состоит из нескольких связанных друг с другом накопителей (хранилищ) информации, которые позволяют пользователям обмениваться и искать имеющуюся там информацию. На рисунке 1.2 показана архитектура NAS.

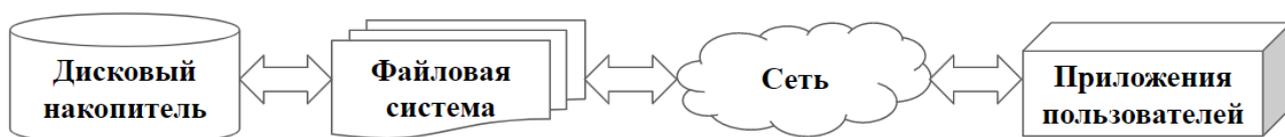


Рис.1.2. Типовая архитектура NAS

Учитывая большой объем информации, создаваемой каждый день по всему миру, неудивительно, что организации ищут эффективные и недорогие устройства для хранения данных, поскольку теперь сетевые хранилища данных могут добавлять новые диски по мере необходимости. Тем не менее, очевидно, что с появлением облачных технологий, анализа больших данных и Интернета вещей, эры доступа к данным в реальном времени, необходим новый подход как для оптимизации данных в хранилищах данных, так и для доступа к этим данным.

Современные облачные технологии для поддержки требований к хранилищу больших данных и программному обеспечению предоставляют пользователям оптимизацию хранилища, безопасность, гибкие методы доставки и расширенную инфраструктуру.

Облака могут содержать не только большой объем необработанных данных, но и данные исходного формата. Новые технологии позволяют обрабатывать их по мере необходимости. Например, Hadoop, созданный с помощью языка Java, позволяет аналитикам хранить большие объемы данных, размещая их на многих недорогих серверах, а затем координировать, интегрировать и обрабатывать данные с помощью MapReduce на виртуальной машине Java (JVM) [22].

В связи с вышеизложенным следует отметить совместный проект проекта Cloud 4 Big Data из мадридского научно-исследовательского института MEDIA Networks, Мадридского политехнического университета и Университета короля Хуана Карлоса [23]. Проект направлен на интеграцию и интеграцию дифференцированных и специализированных технологий на единую, унифицированную платформу, учитывает требования облачных технологий больших данных, «Интернета вещей» и «смарт» – технологий и позволяет добиться в электронном образовании: гибкости, масштабируемости, доступности, качества обслуживания, простоты использования, безопасности и конфиденциальности.

Другими проектами технологии Hadoop являются приложения Spark Apache и Apache Storm [24]. Эти приложения позволяют легко и надежно обрабатывать неограниченные потоки данных в реальном времени в параллельном режиме, постепенно увеличивать инфраструктуру и сохранять данные без потерь. Spark-проект лаборатории Университета Беркли (2009, США), используется для распределенной обработки неструктурированных и полуструктурированных данных. Он состоит из ядра и нескольких приложений для обработки запросов (Sparql SQL), настроек для потоковой распределенной обработки данных (Spark Streaming), набора библиотек машинного обучения (Spark Malibu), программного обеспечения для распределенной обработки графики (Graph).

Помимо Apache Spark и Storm, для разработчиков Java интересен еще один проект- DeepLearning4J [25], предназначенный для создания библиотеки машинного обучения, интегрированной с Hadoop и Spark для Java и Scala, который может импортировать данные из сети и создавать многослойную нейронную сеть, использовать алгоритмы обучения с учителем и без учителя для решения проблем. Этот проект можно использовать для решения следующих задач электронного обучения: распознавание лица или изображения, голосовой поиск, Распознавание речи (превращение ее в текст), регрессионный анализ данных и др.

Осы міндеттердің барлығы MapReduce, NoSQL, Hadoop және басқалары сияқты заманауи технологиялар мен бағдарламалық құралдарды қолдана отырып, үлкен көлемдегі деректерді өңдеуді қажет етеді.

1.3. Выявление ситуаций, связанных с большими данными при создании цифрового сервиса прогнозного анализа качества образования.

Анализ литературных источников позволяет выделить следующие основные группы методов прогнозирования успеваемости студентов:

- методы, основанные на регрессионных моделях;

- методы кластерного анализа;
- методы, основанные на дискриминантных моделях.

Методы, основанные на регрессионных моделях, например, с использованием многофакторной линейной регрессионной модели, позволяют выявить связь между знаниями и навыками учащихся по общепрофессиональным и специальным дисциплинам, связанным с поддерживаемыми курсами.

Методы кластерного анализа позволяют разделить объекты по всей совокупности признаков, которые могут быть исходным уровнем знаний учащихся, уровнем сформированных компетенций, количеством пропусков занятий. Справочные значения параметров для каждого кластера определяются как средние данные для каждой типологической группы студентов. В качестве центров будущих кластеров используются референсные значения, вокруг которых группируются наиболее близкие по значениям выбранных параметров объекты.

Методы, основанные на дискриминантных моделях, предполагают, что они должны заранее определить, какие факторные признаки могут повлиять на успеваемость студентов, и использовать эти факторы для классификации студентов по успеваемости, а также для улучшения качества подготовки путем корректировки основных признаков. Дискриминантная модель оптимально распределяет множество объектов на подмножества и осуществляет классификацию новых объектов при условии, что заранее неизвестно, к каким классам они относятся.

Однако зачастую предположения, полученные формальными методами, неэффективны, так как поведение людей, в частности студентов, нелогично. В последние годы наибольший интерес представляют искусственные нейронные сети, представляющие собой математические модели, основанные на принципе организации и функционирования биологических нейронных сетей. Нейронные сети не запрограммированы в обычном смысле этого слова, они обучены. В

процессе обучения нейронная сеть может обнаруживать сложные связи между входом и выходом, а также выполнять обобщения. После считывания сеть может предсказать будущее значение определенной последовательности, основываясь на нескольких предыдущих значениях. Следует отметить, что прогнозирование возможно только тогда, когда предшествующие изменения в определенной степени предопределяют будущее [43].

Технологии обработки больших данных и интеллектуального анализа (Data Mining) используются для создания прогнозов будущего путем оценки имеющихся и текущих данных.

Data Mining-это:

- совокупность методов поиска данных, позволяющих выявить в накопленных данных ранее неизвестные отношения, практические полезные и доступные знания, необходимые для принятия оптимальных решений в различных сферах человеческой деятельности [36];

- многопрофильная отрасль, возникшая и развивающаяся на основе прикладной статистики, распознавания образов, искусственного интеллекта, машинного обучения, теории баз данных, алгоритмизации, визуализации и других дисциплин [37].

Существует четыре типа анализа Data Mining: описательный, диагностический, прогностический и инструктивный, которые отличаются друг от друга.

Описательная аналитика (Descriptive Analytics) использует имеющиеся данные для выявления причин определенных событий; обобщает исходные результаты, полученные в результате наблюдений и экспериментов; собирает данные в удобной для анализа и восприятия форме; использует основные инструменты аналитики для создания отчетов (например, инструменты, доступные в Microsoft Excel). Описательная аналитика- это основной вид

аналитики, который называется «Что случилось?» и отвечают на вопросы. Основная цель описательной аналитики - получение информации о событиях, которые произошли для дальнейшего изучения и использования.

Диагностическая аналитика-это усовершенствованная форма аналитики, которая позволяет детализировать данные для определения причины конкретного события, устанавливая причинно- следственные связи. В этом случае используются статистические методы анализа данных и средства их визуализации. Диагностическая аналитика отвечают на вопросы «почему это произошло?».

Прогнозная аналитика-один из самых популярных и широко используемых видов аналитики. Он решает проблемы, связанные с прогнозированием и прогнозированием на больших выборках данных, и отвечают на вопросы «что может произойти?». Для построения будущих прогнозов используются методы математического моделирования, теории игр, производства данных, машинного обучения и др. Прогностические модели дают полную картину данных, отражают связь между различными факторами для дальнейшего анализа и оценки рисков и возможностей.

Предписывающая аналитика (Prescriptive Analytics) использует результаты описательного, диагностического и прогностического анализа, анализирует все собранные и обработанные данные для поиска наилучшего решения, задается вопросом и отвечают на вопросы «Что делать?». Этот тип аналитики позволяет фиксировать решение конкретной задачи, постоянно отслеживая результаты и оптимизируя процесс принятия решений.

В зависимости от поставленных задач используется тот или иной вид анализа данных или несколько из них [38]. Кроме того, для проведения анализа определенного вида используется собственный набор методов.

Таким образом, основными методами прогнозной аналитики являются методы, основанные на статистике и машинном обучении. К ним относятся [39]:

- классификация (алгоритмы: наивный Байес, деревья решений, K-близкие соседи, опорные векторные машины и т.д.);

- регрессия (линейная или полиномиальная регрессия);

- кластеризация (метод k-среды, mean-shift, fuzzy c-means, DBSCAN, Agglomerative и др.);

- уменьшение размера (метод упаковки, фильтрация, основные компоненты, SNE и т. д.).

Широко применяются методы прогнозного анализа в различных сферах деятельности.

Для снижения текучести кадров в области управления персоналом крупной инвестиционно-банковской компании Credit Suisse [40] внедрена система, основанная на методах прогнозного анализа. Он анализирует возможные причины, по которым сотрудники планируют увольнение, и информирует высшее руководство о намерениях в дальнейшем принимать кадровые решения.

В области маркетинга американская компания Amazon [41] использует алгоритм, который предоставляет потенциальному покупателю информацию о аналогичных продуктах на основе предыдущих покупок или просматриваемых страниц продукта.

Учебная аналитика в сфере образования позволяет собирать, представлять, анализировать, измерять данные о студентах, предметной области, технологиях и методах обучения с целью повышения качества образовательного процесса. Для прогнозирования успешного завершения обучения в старших классах и готовности учащегося к поступлению в вуз можно использовать прогностическую аналитику. Использование учебной аналитики студентом помогает снизить риск

отчисления, способствует построению индивидуальных образовательных траекторий и улучшению образовательных результатов.

Для эффективного решения прогнозных задач необходимо разработать и построить прогнозную модель на основе следующего алгоритма.

1-этап. Планирование. На этом этапе необходимо определить цели и задачи модели, описать предметную область, определить область и охват применения и набор используемых данных. Также необходимо сформулировать вопросы, на которые отвечает модель, например, что предсказать? Какие большие данные нужны? Что такое источник данных?

2-этап. Сбор данных для моделирования. Необходимые данные могут быть получены как из внутренних источников (данные клиентов, студентов, транзакций и т. д.), так и из внешних источников (данные из опросов, исследований, открытых библиотек и т. д.).

3-этап. Анализ данных и статистика. На этом этапе подготавливаются данные для обучения нейронной сети. Анализ данных-это процесс проверки, очистки, исследования, преобразования, фильтрации и моделирования для выявления полезной информации и выводов. Статистический анализ позволяет проверить прогноз на основе применения стандартных статических моделей. Для лучшего понимания и определения того, какие элементы данных влияют на результаты прогноза, необходимо провести разведывательный анализ данных (EDA). Это помогает правильно очистить данные и получить признаки, необходимые для моделирования [42].

4-этап. Моделирование и построение модели. Перед созданием модели прогнозирования необходимо выбрать метод машинного обучения, определить тип нейронной сети и тип модели, настроить параметры и гиперпараметры нейронной сети (Количество слоев и нейронов сети, архитектура нейронной сети, количество эпох обучения и т.д.), выбрать алгоритм обучения нейронной сети.

При построении модели необходимо использовать методы перекрестной проверки данных для повышения ее эффективности. Наиболее распространенные методы: тренировочный тест, тренировка / проверка / тестовый образец, LOOCV (leave-one-out cross-validation (кросс валидация, которая оставляет друг друга)), K-fold cross-validation (k-кратная кросс валидация).

5-этап. Оценка и контроль модели. На этом этапе созданная модель проверяется на надежность и точность, оценивается качество обучения сети, изменяются и выбираются гиперпараметры сети. В этом случае нейронная сеть переучивается с новыми значениями гиперпараметров. Необходимо оценить качество данных в тестовом наборе до тех пор, пока качество сети не будет соответствовать требованиям прогнозируемой модели.

6-этап. Прогнозирование и позиционирование модели. На заключительном этапе окончательные отчеты составляются и представляются в подробном, простом в использовании, согласованном формате, что позволяет быстро принимать и стратегически использовать результаты прогнозирования. Построенная модель представляет собой вероятностное предположение, основанное на имеющихся данных. Развертывание-это размещение последней модели в среде, которую могут использовать другие приложения.

Для оценки успешности обучения используются следующие методы машинного обучения: линейная регрессия, логистическая регрессия, глубокие нейронные сети и дерево решений.

Проанализировав методы обработки данных, можно отметить следующие преимущества нейронных сетей перед традиционными вычислительными системами:

- нейронная сеть, способная решать задачи, в которых неизвестны закономерности развития ситуации и связи между входными и выходными данными;

- возможность работы при большом количестве неинформативных, шумных входных сигналов;

- нейронная сеть способна адаптироваться к изменениям окружающей среды. То есть нейронные сети, обученные в определенной среде, могут быть легко перепрофилированы для работы в условиях незначительных колебаний параметров среды;

- нейронные сети обладают потенциальной сверхвысокой производительностью за счет использования массового параллелизма обработки информации;

Преимущества использования дерева решений:

- процесс скоротчения;

- формирование правил в областях, где эксперту трудно формализовать свои знания;

- выведение правил на естественном языке;

- модель интуитивной классификации;

- высокая точность прогнозирования, сравнимая с другими методами (статистика, нейронные сети);

- построение непараметрических моделей.

По этим и многим другим причинам методология дерева решений является важным инструментом в работе специалиста аналитика. В результате исследования было установлено, что алгоритм деревьев решений подходит для решения этой задачи [44].

2. МОДЕЛИ ПРОГНОЗНОГО АНАЛИЗА КАЧЕСТВА ОБРАЗОВАНИЯ В УНИВЕРСИТЕТЕ НА ОСНОВЕ ТЕХНОЛОГИЙ BIG DATA

2.1. Алгоритмы оценки участия студентов в учебном процессе

Уровень успеваемости студента в вузе является формой диагностики и прогнозирования степени будущего специалиста. Успехи студентов - это показатель деятельности вуза в решении учебно-воспитательных задач. Для максимально эффективного решения этих задач необходима постоянная объективная оценка, коррекция и управление. Однако управление невозможно без прогноза. Поэтому возникает необходимость прогнозирования успеваемости студентов на всех этапах обучения.

Стремительное развитие информационных технологий позволяет автоматизировать многие сферы человеческой деятельности, повысить их эффективность. Возможности этих технологий могут быть использованы и в сфере образования.

В данном исследовании уделяется внимание построению прогнозной модели успеваемости студентов на основе текущих оценок с использованием технологий анализа данных. Мерой измерения качества знаний, полученных для конкретного студента, являются его оценки по пройденным дисциплинам. Своевременные меры по оказанию помощи студентам, не выдерживающим учебной нагрузки, являются одним из основных разделов по учебно-воспитательной работе в вузе, которые влияют на показатели качества образования в учебном заведении. На самом деле, на успеваемость студента влияет множество факторов, и одним из наиболее важных является мотивация к обучению, моральная ситуация, отношения с сокурсниками и т. система определяет проблемных студентов и позволяет детально работать с каждым и определяет, какие проблемы у этого студента.

Далее рассмотрим алгоритмы машинного обучения, используемые для решения аналогичных задач в области прогнозирования успеваемости студентов. Приводятся наиболее распространенные методы и алгоритмы, а также описание существующей работы с конкретными задачами, для которых эти методы используются.

Некоторые источники используют метод кластеризации, что, на наш взгляд, совершенно неверно, так как у нас есть признаки класса, а именно количество заимствований, поэтому в данном случае необходимо использовать методы классификации. Кластеризация - это методы машинного обучения без учителя, а классификация и регрессия, в свою очередь, обучение с учителем, поскольку их исходные данные являются маркерами для каждого класса. В исследовании

использовались методы кластерного анализа для оценки итоговой оценки по дисциплине [45].

К-алгоритм ближайших соседей

Одним из наиболее распространенных способов решения подобной задачи является алгоритм k - ближайших соседей:

К- алгоритм ближайших соседей (KNN, k nearest neighbors) - это тип управляемого алгоритма машинного обучения, который можно использовать как для классификации задач регрессионного прогнозирования, так и для прогнозирования. Этот алгоритм прост в понимании, но тем не менее он доказал свою эффективность в ряде задач и используется не только в учебных целях. Алгоритм ближайших соседей еще называют ленивым классификатором, так как в процессе обучения он не создает никакой модели, а только хранит данные. Все вычисления начинаются только тогда, когда необходимо классифицировать новые данные.

Суть этого алгоритма заключается в том, что прогнозирование новых значений данных основано на их близости к данным, уже установленным в учебном наборе. Другими словами, если между ближайшими соседями новой точки данных находятся 4 точки класса А и 1 точка класса В, эта новая точка определяется как класс А. Таким образом, k -ближайшие соседи будут 2 наиболее важных параметра в алгоритме. В частности, метрика расстояния (Евклидово, манхэттенское или Хаммингское) и количество соседей, которые мы рассматриваем. Наглядное представление работы алгоритма представлено на рис.2.1.

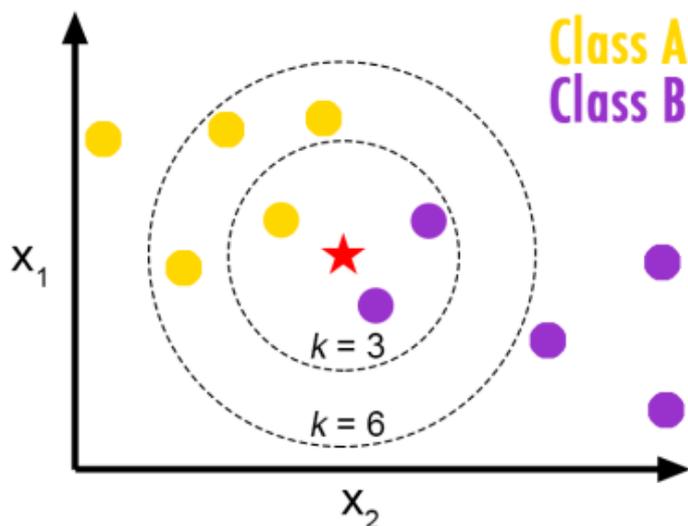


Рис.2.1. Работа алгоритма KNN

На рисунке 2.1 показан массив исходных точек, закодированных цветом в зависимости от классов. Звездочка отмечает точку, которую нужно классифицировать. Все точки даны двумерно по осям X_1 и X_2 . Если обозначенное число ближайших соседей равно 3, немаркированный объект относится к классу B, если 6, то к классу A.

Этот алгоритм имеет много нюансов, например, мы можем добавить вес к данным во время голосования из-за близости к немаркированному объекту. Именно эти нюансы делают алгоритм KNN актуальным при решении многих задач.

Преимущества алгоритма:

- Простота понимания и интерпретации;
- Хорошо работает на нелинейных данных;
- Соответствует задачам универсального алгоритма, классификации и регрессии;
- Имеет относительно высокую точность.

Недостатки алгоритма:

- В отличие от алгоритмов, использующих построение моделей, большие потери памяти для хранения всех данных;

- Чувствительность к масштабу данных;
- Медленное прогнозирование в случае большого объема данных;
- Повышенная чувствительность к «шуму» в данных.

В статье А.А. Будаевой «Прогнозирование индивидуальной успеваемости студентов в вузе» [46] этот алгоритм используется для классификации оценок отдельного студента по каждой дисциплине на основе оценок его прошлых оценок и оценок студентов прошлых курсов с наиболее сходными параметрами по этим дисциплинам. В данной статье приведена высокая точность алгоритма для данной задачи, максимальная погрешность оценки прогноза составила 0,55 балла.

Этот алгоритм также используется в статье «Student performance prediction using support vector machine and k nearest neighbor» [47], в которой предполагается оценка экзамена по определенной дисциплине на основе оценок студента по предыдущей дисциплине, его участия и оценок рубежного контроля.

Метод опорных векторов

Опорный векторный метод (SVM- support vector machines) - набор аналогичных алгоритмов типа «обучение с учителем», используемых для решения задач классификации и регрессии. Этот метод является одним из самых популярных методов обучения и относится к группе линейных классификаторов. Отличительным свойством метода опорного вектора является непрерывное уменьшение эмпирической ошибки классификации и увеличение разрыва. Поэтому этот метод также известен как метод классификатора с максимальным разрывом.

Основную идею метода опорных векторов можно показать на примере ниже: на плоскости есть точки, обозначенные 2 классами, которые линейно делятся. В этом случае полученная функция будет плоскостью, разделяющей эти классы. Тем не менее, можно построить много гиперплоскостей, которые разделяют

эти классы. Чтобы найти оптимальную гиперплазию, необходимо найти максимальную сумму нормальных векторов класса А и класса В. Визуальное представление этого метода можно увидеть на рис. 2.2. На этом рисунке опорные векторы перпендикулярны нормам, приведенным на рисунке.

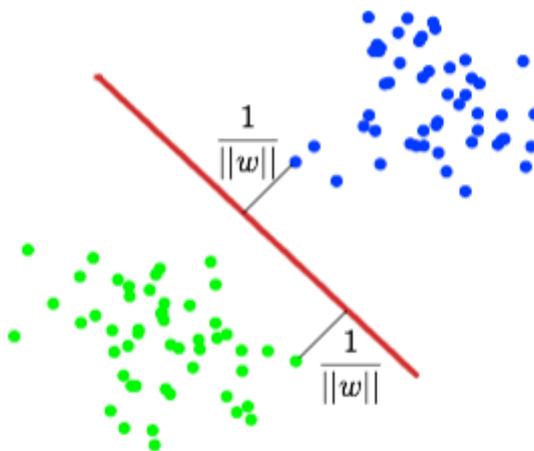


Рис.2.2. Метод опорных векторов

Формальное описание этого метода выглядит следующим образом: имеем $\{(X_1, C_1), (X_2, C_2), \dots, (X_i, C_i)\}$ пусть модель обучения,

где: X_i -реальный вектор размера p

C_i -значение 1 или -1, принимаемое классом

Опорный векторный метод строит классификационную функцию следующим образом:

$F(X) = \text{sign}([w, x] + b)$, (1), где $[,]$ - скалярное произведение

w -разделительный нормальный вектор гиперплазии

b -вспомогательный параметр

Таким образом, все это можно записать как проблему оптимизации, которая имеет решение и в то же время является единственной.

$$\begin{cases} \|\mathbf{w}\|^2 \rightarrow \min \\ c_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad 1 \leq i \leq n. \end{cases} \quad (2.1)$$

Эта задача решается методом квадратичного программирования и с помощью множителей Лагранжа.

Рассматривалась ситуация, когда было два разделенных класса. Практически всегда классы не делятся линейно и стоит задача классификации более двух классов. Чтобы решить задачу с линейными неделимыми классами, мы позволяем классификатору делать ошибки в модели обучения. Для такого предположения запишем уравнения.

$$\begin{cases} \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \rightarrow \min_{w,b,\xi_i} \\ c_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i, \quad 1 \leq i \leq n \\ \xi_i \geq 0, \quad 1 \leq i \leq n \end{cases} \quad (2.2)$$

Где С-параметр установки метода,

ξ_i -значение допустимой ошибки.

Для решения многоклассовых задач используется обобщенный метод опорных векторов, так как переход к классификации на многие классы осуществляется путем деления на 2 класса, как на приемлемый класс и несоответствующий. Эта стратегия также называется «Против всех» и используется для использования двоичных классификаторов для многоклассовых задач.

Преимущества алгоритма:

- Проблема хорошо изучена и имеет одно решение;
- Принцип оптимальной разделительной гиперплоскости приводит к надежной классификации;

- Количество нейронов в скрытом слое равно двойной слоистой нейронной сети, которая автоматически определяется как число поддерживающих векторов.

Недостатки алгоритма:

- Нестабильность шума, экстремальные значения в исходных данных напрямую влияют на построение разделительной гиперплоскости.
- Нет выбора функции.
- Для каждой задачи необходимо отдельно выбрать методы создания ядер и исправления пробелов.

Этот алгоритм используется в статье «Student performance prediction using support vector machine and k nearest neighbor» [47], в которой предполагается оценка студента за экзамен по определенной дисциплине, его оценки по предыдущим предметам, посещаемость и промежуточные контрольные оценки.

Нейронные сети

Помимо вышеперечисленных методов, нейронные сети также используются для прогнозирования успеваемости студентов. Существует много информации о возможности использования нейронных сетей для решения этой проблемы, но нет информации о фактической реализации таких прогностических моделей.

Нейронные сети - это математические модели, основанные на принципе организации и функционирования биологических нейронных сетей. В обычном смысле этого слова нейронные сети не запрограммированы, они обучены. В процессе обучения нейронная сеть может обнаруживать сложные связи между входом и выходом, а также выполнять обобщения. После считывания сеть может предсказать будущее значение определенной последовательности, основываясь на нескольких предыдущих значениях.

Иллюстрация принципа функционирования нейронной сети представлена на рис. 2.3. Нейронная сеть состоит из нейронов, слоев и синапсов. В качестве

цвета с изображением различных узлов нейронов. Все узлы одного цвета относятся к одному слою нейронной сети. Синапсы- это линии, которые соединяют нейроны в одном слое с нейронами в другом слое. Синапсы имеют только один параметр, это вес.

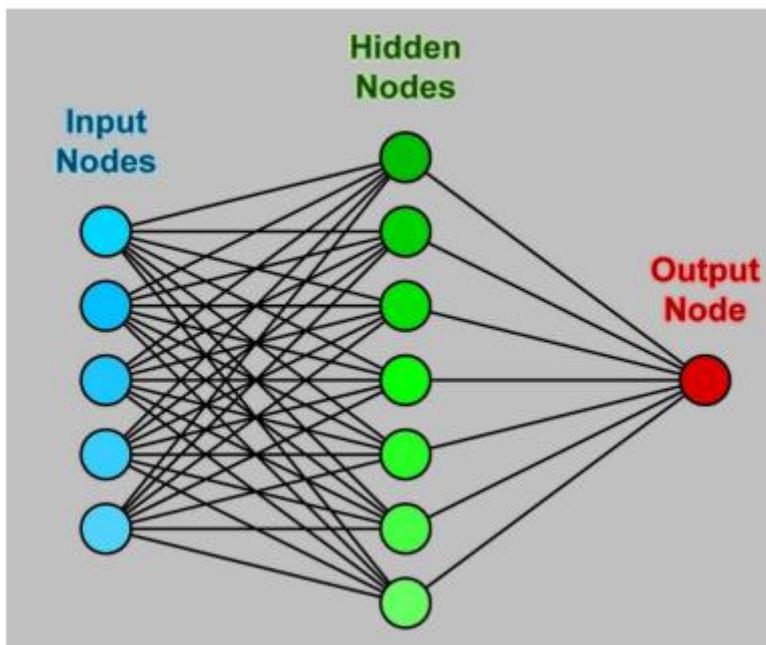


Рис.2.3. Нейронные сети

Каждый нейрон выполняет определенную математическую функцию, поэтому в него входит множество значений, а на выходе один. Таким образом, на выходе обученная нейронная сеть приобретает определенное значение, данное ей.

Преимущества:

- Соппротивление шума вводимых данных;
- Самообразование и творчество. Возможность решения нерешенных задач другими алгоритмами;
- Адаптация к изменениям, переобучение.

Недостатки:

- Для больших сетей невозможно заранее оценить время обучения сети;
- Сложность интерпретации результатов;

Рассмотрим подробнее использование нейронной сети для решения задачи прогнозирования успеваемости студентов. Ясинский И.Ф. В статье «Опыт прогнозирования успеваемости студентов с помощью технологии нейронной сети» [48] модель нейронной сети обучена прогнозировать, будет ли студент перспективным или нет.

В статье «Анализ и прогнозирование успеваемости студентов на основе радиальной базовой нейронной сети» [49] нейронная сеть используется для прогнозирования оценок в курсе информатики.

В статье «Прогнозирование успеваемости первокурсников по результатам сдачи Единого государственного экзамена» Харламова И.Ю. [50] рассмотрела проблему классификации студентов по результатам ЕГЭ.

В этом разделе были рассмотрены наиболее часто используемые алгоритмы для решения задачи прогнозирования успеваемости студентов на основе различных исходных данных и для решения различных задач, будь то успеваемость по определенному предмету или общее представление успеваемости по всем предметам. Рассмотренные примеры прогнозирования не носят систематического характера, а лишь являются попытками приблизиться к решению данной задачи. Очевидно, что для успешного решения проблемы необходимо использовать несколько методов и сравнить их результаты.

Качество моделей машинного обучения сильно зависит от качества входных данных. Тем не менее, фактические данные очень часто плохо структурированы, несуществующие значения, шум и неправильные значения. При неправильной подготовке данных никакие настройки алгоритмов машинного обучения не могут обеспечить высокую предсказуемую точность этих моделей. Подготовка данных перед анализом занимает около 80 процентов времени специалиста машинного обучения, но это необходимая работа.

В данной работе для подготовки данных используется методика, описанная в статье Е.И. Губина «Методика подготовки больших данных для прогностического анализа» [51].

Для ознакомления с данными и подготовки их к дальнейшему анализу используется язык программирования Python и его библиотеки. Выбор языка программирования Python обусловлен высокой производительностью в обработке данных, простотой и большим количеством библиотек для машинного обучения. Python является одним из лучших языков программирования для работы с данными. В этой исследовательской работе используются следующие библиотеки Python:

- **NumPy.** Эта библиотека добавляет поддержку больших многомерных массивов и матриц, а также добавляет команды высокого уровня для математических функций с очень высокой производительностью в этих массивах [52].
- **Pandas.** Программная библиотека для обработки и анализа данных.
- **telebot** (pyTelegramBotAPI). Хорошая и легкая библиотека для создания бота Telegram на языке Python.
- **aiogram.** Очень простая и полная асинхронная структура для API Telegram Bot, написанная на Python 3.7 с использованием Asyncio и aiohttp. Это поможет сделать ботов быстрее и проще.
- **python-telegram-bot.** Предоставляет интерфейс Python для TelegramBotAPI. Он совместим с версиями Python 3.6+.
- **Telepot.** Помогает создавать приложения для TelegramBotAPI. Он работает на Python 2.7 и Python 3. Для Python 3.5+ у него есть версия синхронизации на основе asyncio.

2.2. Модели прогнозирования учебной траектории студентов

В настоящее время актуальной тенденцией в высшем образовании является персонализация обучения. В результате возникает необходимость формирования

индивидуальной образовательной траектории при активном участии обучающихся.

В связи с этим одним из наиболее перспективных направлений повышения эффективности образования является проектирование и институциональное сопровождение индивидуальных образовательных траекторий обучающихся. В основе индивидуальной образовательной траектории лежит осознанный и ответственный выбор целевого направления реализации своего профессионального образовательного потенциала в соответствии со способностями субъекта. В этом случае студент выступает субъектом проектирования своей профессиональной биографии.

«Индивидуальная образовательная траектория-это индивидуальный путь реализации личностного потенциала каждого учащегося в образовании». Из определения следует, что путь человека в образовании определяется не только логикой предметов и областей знаний, но и личностным потенциалом ученика, то есть его склонностями и способностями [61].

При определении траектории сегодня можно выбрать:

- уровень освоения профессиональной образовательной программы;
- предметная область, включающая разнообразие программ дисциплин и выбор дополнительных образовательных программ, включая профессиональные модули вариативной части;
- методы и формы обучения, формы самостоятельной работы студентов;
- формы контроля результатов обучения;
- темпов обучения; - количество и содержание профессиональных тестов, предлагаемых в рамках обучения (участие в соревнованиях, олимпиадах и т.д.);
- место прохождения стажировки;
- темы научно-исследовательских и проектных работ;
- тема выпускной квалификационной работы и др.

Эффективность личностного развития определяется возможностью всестороннего учета индивидуальных особенностей обучающихся и создания условий для их развития на каждом этапе обучения. Поэтому в современном университете актуальным является решение проблемы проектирования индивидуальной траектории профессионального развития студентов [53].

Изучение данной проблемы осуществляется рядом авторов. В данной исследовательской работе мы рассмотрим основные достижения, доступные в настоящее время, которые, на наш взгляд, позволяют внедрить индивидуальные траектории развития студентов на начальном этапе обучения в вузе, при изучении математических и естественнонаучных дисциплин и собственно в общую индивидуальную траекторию.

В психологии планирование траекторий профессионального развития и закономерности его реализации связаны с профессиональным самоопределением Э.Ф. Зеера и Э.Э. Семенюка, прогнозированием профессионального будущего человека на протяжении всей его жизни [54]. В исследованиях Э.Ф. Зеера [55] делается акцент на необходимости развития способностей личности, самостоятельной оценки их профессиональных склонностей, формирования знаний о возможности трудоустройства, способах карьерного роста, конкурентоспособности.

В дидактике И.Ф. Бережная разрабатывает проблему подготовки студентов университета на основе проектирования индивидуальной траектории профессионального развития [53]. Решение, предложенное этим автором, строится на основе закона развития систем, концептуально адаптированных к чтению С.М. Годника.

Проблема реализации (проектирования) индивидуальных образовательных траекторий широко освещена в современной литературе, в разработке которой приняли участие Е.А. Александрова, Г.А. Бордовский, Э.Ф. Зеер, Е.А. Климов,

А.М. Маскаева, Н.Н. Суртаева, Э.Э. Симанюк, А.В. Хуторский, Г.П. Щедровицкий, И.С. Якиманская и др. ученые [53-59].

Индивидуализация учебного процесса делает осмысленной учебную деятельность через выбор определенного вида деятельности, включение в учебный процесс личностных смыслов, а также формирование собственного образовательного заказа и их отношения к образовательным перспективам. Видение перспектив имеет решающее значение для проектирования и реализации индивидуальных образовательных траекторий студентов в условиях непрерывного образования [54].

С точки зрения Т.М. Ковалевой, индивидуальная образовательная траектория имеет пространственно-временную характеристику. Именно траектория индивидуального образовательного движения, «след» линии движения ученика, сформированной путем закрепления содержания его испытаний и опыта, особенностей учебных достижений и индивидуального образовательного пространства, позволяет реализовать этот педагогический прогноз и тьюторский проект [62].

Под индивидуальной образовательной траекторией студента мы понимаем личностный путь, направленный на реализацию личностных устремлений в образовании, разработку жизненных стратегий, формирование основ личностного творческого и профессионального развития личности студента, которые самостоятельно создаются и реализуются субъектом образовательного процесса при осуществлении педагогической поддержки наставника в самоопределении и самореализации.

Еще одно понятие, тесно связанное с понятием «большие данные», — «цифровой след», под которым понимается определенный набор действий, совершаемых пользователем во всемирной паутине или с использованием других цифровых носителей. Синонимами этого понятия являются «электронный след», «интернет-след», «кибер-тень» и другие. Все это цифровая информация, которую

человек оставляет после работы с электронными устройствами и хранится в виде файлов cookie. Цифровой след может оставить как один человек, так и группа людей. Если человек не знает цифрового следа, то такой след называется пассивным. Например, студент ищет информацию на различных веб-сайтах, а система хранит определенную информацию, поэтому через простые поисковые системы можно легко собрать большой объем информации о конкретном студенте. С другой стороны, мы сознательно оставляем много информации о себе, размещая фотографии и сообщения в социальных сетях, общаясь на форумах, публикуя свои личные данные в Интернете. Это называется активным цифровым следом, который можно использовать для сбора данных о конкретном человеке или группе людей.

Каталин Фехер, анализируя проблему цифрового следа: универсальные модели частных онлайн - стратегий в основном подчиняются сознательным решениям, в результате чего пользователи контролируют свои цифровые следы на 70%. Тем не менее, оставшиеся в Интернете 30% действий совершаются пользователями неосознанно" [63].

В настоящее время цифровой след активно используется в образовании. В сфере образования цифровым следом являются письменные работы учащегося, конспекты, тесты, онлайн-курсы, фотографии и т. д. Современные технологии позволяют распознавать лицо, голос, переводить речь в текст и наоборот за считанные секунды. На основе анализа и специальной обработки этой тропы участникам могут быть даны некоторые предложения, рекомендации и персонализированные тренировки. Цифровой след позволяет образовательным организациям лучше понимать поведение учащихся, оказывать им необходимую поддержку и быть наставниками в развитии способностей студенческой аудитории.

Все современные информационно-коммуникационные технологии и основанные на них методы обучения должны быть направлены на обеспечение

возможности формирования каждого студента как уникального специалиста, прошедшего собственную траекторию становления и развития, накопившего собственные компетенции.

В современной науке предпринимаются попытки найти эффективные пути применения цифровых технологий в образовательной среде. Уже известны некоторые особенности применения искусственного интеллекта и дополненной реальности в реальной педагогической практике. Сегодня существует множество инструментов для анализа и систематизации больших данных. Поэтому одной из целей данного исследования является анализ опыта цифровизации образовательного пространства в разных странах за последнее десятилетие.

В настоящее время имеется достаточно научных публикаций по использованию больших данных и цифрового следа в образовании. Анализ научных источников показывает, что точное определение понятия «цифровой след» в образовании еще не выработано исследователями, по разным взглядам к нему относят информацию, написанную в журналах цифровой системы, а также «цифровые артефакты», которые учащиеся помещают в цифровую среду. Ниже приведены работы зарубежных авторов оценить потенциал этих ресурсов и использовать их в дидактических целях.

Структурно-информационные характеристики цифрового образовательного пространства исследованы Е.Г. Галимовой, А.В. Коньшевой, О.А. Калугиной, З.М. Сизовой как методы оценки результатов учебно-познавательной деятельности студентов в процессе обучения математике [64].

На основе анализа данных, относящихся к цифровому следу, Д. Аскона, И. Сяо и А.Ф. Смитон разработали собственную методику, позволяющую автоматически выявлять «группы риска» среди студентов в аспекте невозможности выполнения заданий по модулям (курсам) компьютерного программирования и при этом сохранять адаптивную обратную связь. Метод реализован в рамках совместного проекта ирландских и американских ученых и

описан в статье «выявление студентов, относящихся к группе риска на уроках компьютерного программирования на основе анализа цифрового следа студентов» [65].

На сегодняшний день большая часть работы по образовательной аналитике и поиску данных образовательной среды ориентирована на онлайн-курсы. Некоторые отечественные и зарубежные исследователи для решения проблем повышения качества и индивидуализации образования проводят экспертизу цифровых жестов, формируемых студентами в образовательной среде, созданной на различных платформах: LMS Moodle, мессенджере WhatsApp, социальных сетях. Специализированные сервисы для мониторинга учебной активности учащихся в системе Moodle позволяют собирать, хранить и интерпретировать различные данные для определения индивидуального пути обучения, адаптации к окружающей среде и корректировки методов обучения (Сухонен, 2019) [66].

И.Й.Л. Нг, Н. Лоу и А.Х.К. Йуен рассматривают цифровой след, оставленный студентами в социальных сетях. Исследователи рассматривают широкое использование студентами цифровых технологий, особенно вне структурированного учебного времени, и отмечают растущий интерес к пониманию студентов (их поведения) биографическими и этнографическими методами [67].

В своей статье В. В. Мантуленко обобщает подходы к использованию цифрового следа студента на высоком уровне, а именно обеспечение преемственности и интеграции уровней образования, организация учебного процесса (например, построение индивидуальных образовательных траекторий), управление системой образования [68].

Работа В. Н. Курбацкого рассматривает цифровое образование с точки зрения образования и творчества студента, что позволяет исследовать современные компетенции студента с учетом его индивидуальных способностей и потребностей [69].

Творческий коллектив под руководством И.Г. Захаровой изучает особенности постановки и решения управленческих задач на основе фундаментального процесса анализа данных цифрового следа студента в образовательной среде вуза, описывает основные виды информационных услуг, обеспечивающих принятие решений на основе данных [70].

Изучение цифрового следа оставленного студента в социальных сетях позволяет исследовать его «цифровую идентичность». Современных студентов бакалавриата, магистратуры можно считать квалифицированными в области цифровых технологий, но развитие их ответственности в процессе создания интернет-контента остается актуальным (Gamacho, Minelli, Grosseck, 2012) [71].

Одним из направлений использования цифрового следа в организации образовательного процесса является образовательная аналитика («учебная аналитика» или «образование на основе больших данных»). Это консолидация, обработка и анализ цифровых следов с целью понимания и оптимизации образовательного процесса и образовательной среды. Он используется для создания индивидуальной образовательной траектории, для совершенствования преподавания и обучения в высшей школе, для улучшения образовательных результатов.

Интеллектуальный анализ цифрового следа знаний учащегося позволяет:

- Пожизненное моделирование профилей образовательных интересов личности;
- Разработка индивидуальных рекомендаций по обновлению и приобретению новых компетенций;
- Разработка рекомендуемых систем навигации по онлайн-ресурсам для создания индивидуальных направлений обучения;
- Мобилизация материалов и продуктов с учетом образовательных потребностей обучающихся при разработке и разработке основных программ образования и обучения;

- Принятие оптимальных управленческих решений по развитию образовательной среды вуза. Изучение цифровых следов учащегося в образовательном пространстве помогает сформировать общее видение новых моделей обучения для всех субъектов образовательного процесса и найти системное решение для их реализации.

2.3. Критерии прогнозного анализа качества образования в университете

Контроль качества знаний студентов позволяет управлять учебным процессом. Для поиска наиболее сложных вопросов необходимо систематически контролировать знания студентов в вузе. Регулярные мониторинговые исследования могут предоставить информацию, необходимую для анализа процесса обучения.

Система контроля не должна зависеть от текущего контроля преподавателя. Необходимо вести постоянный контроль в течение всего учебного года во всех группах по всем дисциплинам в зависимости от охвата учащимися основных тем. Существующими методами мониторинга являются методы анкетирования, анализа и тестирования результатов обучения.

Анкетные методы позволяют получить информацию о развитии учащихся на основе анализа письменных или устных ответов на стандартные специально подобранные вопросы.

При анализе результатов учебной деятельности изучаются письменные тексты, графика, творческая работа учащихся.

Тестирование является одним из объективных методов сбора данных об уровне развития учащихся. Стандартно-ориентированное тестирование позволяет выявить и сравнить результаты, полученные тестом.

Мониторинг качества образования студентов решает следующие вопросы: повышение уровня знаний каждого студента в отдельности; Повышение рейтинга вузов; Привлечение высококвалифицированных специалистов государством.

Система контроля включает текущий, промежуточный и итоговый контроль. Формы контроля и их количество определяются преподавателем, регламентируются кафедрой и деканатом.

Результаты мониторинга можно рассматривать в динамике по сравнению с предыдущим учебным годом. Методическая работа преподавателей и студентов планируется по результатам мониторинга [72].

Образовательная аналитика (LA), которая использует данные учащихся для улучшения преподавания и обучения, позволяет высшему образованию постепенно использовать собранные данные для понимания и улучшения преподавания, чтобы оптимизировать успеваемость учащихся [73].

К основным группам методов образовательной аналитики относятся: методы прогнозирования, методы определения структуры, анализа взаимосвязей, обработки данных и представления их лицам, принимающим решения.

В целях понимания и оптимизации учебного процесса и среды, в которой этот процесс протекает, направление исследований, основанное на сборе, анализе и представлении информации об обучающихся и их действиях на компьютере, теперь определяется термином «учебная аналитика». В настоящее время образовательная аналитика объединяет множество исследований, основанных на использовании цифровых записей деятельности учащихся в образовательных целях. Появление образовательной аналитики связано с растущим количеством данных о деятельности субъектов образования, которые могут собирать компьютеры для дальнейшего использования в образовательном процессе. В тех случаях, когда участие субъекта образования в образовательном процессе осуществляется посредством цифровых устройств, данные о поведении участника могут быть автоматически получены, сохранены и использованы для анализа. Простейшим примером данных, с которыми может работать образовательная аналитика, являются данные веб-аналитики. Например, в лог-файлы веб-сервера записывается информация о том, откуда пришел тот или иной

посетитель, когда и сколько времени он провел на сайте, что он видел и скачивал, какой у него был браузер и какой у него IP-адрес. Благодаря этим записям, содержащим следы действий пользователя, исследователь может получить достоверную информацию и ответы на вопросы:

1. Какой браузер используют студенты для работы с материалами дистанционного обучения?
2. Сколько времени студенты тратят на работу с материалами курса?
3. Какие страницы образовательного ресурса наиболее интересны и часто просматриваются?
4. Какие члены исследовательской группы заинтересованы в материалах?
5. Какие материалы наиболее горячо обсуждаются?
6. Какие члены ассоциации часто создают и редактируют совместные страницы?

Записи, используемые для чтения аналитики, могут включать в себя данные о местоположении учащихся, журналы специальных программ, в которых хранятся хронологические записи, и все действия пользователя. Во всех случаях для обучения аналитике важны следующие особенности:

1. Первичными данными являются цифровые записи действий субъектов образования: время нахождения на страницах, созданные объекты, количество и содержание правок, эмоциональное состояние и т.д. Эти данные являются результатом регулярного контроля и регистрации сигналов о действиях преподавателей и учащихся, а не получены из опросов и тестов.
2. Методы обработки данных, получения и визуализации знаний доступны всем субъектам образования и могут быть использованы для корректировки их деятельности.
3. Знания, полученные в результате анализа, помогают преподавателям и учащимся лучше понять события, происходящие в сценарии обучения – с

чем работают участники, как взаимодействуют, чем занимаются, какие инструменты используют, среду обучения [75].

Для современной системы образования особое значение имеет вопрос эффективного контроля и научно обоснованных критериев и показателей качества образовательных учреждений. В педагогических исследованиях существует несколько основных подходов к этому вопросу.

В справочной и специальной литературе понятие «критерий» обычно трактуется как критерий оценки предмета или явления; признак, полученный в качестве основы для классификации, по которой сравнивают различные явления или процессы. Качественная сформированность, степень наглядности критерия отражается в конкретных показателях, которые являются наиболее узкими категориями по объему и содержанию. Являясь составной частью критерия, показатель отражает типичные специфические свойства и особенности изучаемых процессов или явлений [76].

Одной из задач исследования является определение критериев и показателей эффективного мониторинга качества образовательных услуг в высшей школе на основе новейшего международного опыта.

Очевидно, что объективность ее состояния и показателей эффективности играет важную роль в формировании эффективной системы мониторинга. В результате многолетнего зарубежного опыта сформулированы основные требования к образовательным показателям.

При формулировании критериев и показателей мониторинга качества необходимо ориентироваться прежде всего на принципы системного подхода, разделив показатели на три условные группы: критерии и показатели обеспеченности, эффективности и результативности системы высшего образования как процесс и управление им как системой и как конечным продуктом этого процесса. Эти три группы показателей должны стать ядром

создания эффективных инструментов измерения качества высшего образования, отвечающих требованиям международных стандартов.

3. МЕТОДИЧЕСКИЕ ОСНОВЫ СОЗДАНИЯ И ИСПОЛЬЗОВАНИЯ ЦИФРОВОГО СЕРВИСА ПРОГНОЗНОГО АНАЛИЗА КАЧЕСТВА ОБРАЗОВАНИЯ

3.1. Особенности анализа и использования больших данных при создании цифрового сервиса прогнозного анализа качества образования

Развитие системы электронного обучения и внедрение цифровых технологий в традиционный образовательный процесс приводит к появлению больших информационных массивов. Анализ образовательных данных (Educational Data Mining) – это направление исследований, связанных с

использованием методов интеллектуального анализа данных (Data mining), машинного обучения и статистики информации, производимой образовательными учреждениями. Анализ образовательных данных часто разрабатывает и совершенствует методы обработки образовательных данных, которые имеют несколько уровней иерархии значений, чтобы лучше понять, как студенты учатся и как сделать учебный процесс более эффективным.

Анализ образовательных данных пытается извлечь закономерности из данных, возникающих в процессе обучения. Эти данные могут быть очень объемными и содержать много данных. Таким образом, некоторые системы управления обучением (learning management system) отслеживают информацию о том, когда студент получил доступ к определенному учебному контенту, сколько раз студенты обращались к этому учебному контенту и сколько минут студент отображал на экране компьютера этот учебный контент, в какой последовательности выполнялись тестовые задания и т.д. И другие данные могут содержать гораздо меньше данных. Например, в электронной зачетной книжке студента содержится упорядоченный во времени список курсов, прослушанных конкретным студентом, оценки, полученные им за определенный курс. Анализ образовательных данных использует оба этих типа данных для выявления закономерностей в обучении студентов. Таким образом, изучая данные электронной образовательной среды, можно установить связь между темами обучения, на которые студент обратил внимание в ходе курса, и итоговой оценкой данного студента.

Анализ образовательных данных тесно связан с аналитикой обучения. Цели и задачи исследований в этих двух областях во многом совпадают. Ряд специалистов [77, 78] считают, что анализ образовательных данных для разграничения двух областей направлен на автоматизацию выявления закономерностей в образовательных данных, а аналитика обучения направлена на подготовку данных для человеческого анализа.

Систематическая работа по сбору и изучению образовательных данных, представленных в электронном виде, началась в 1980 – х годах в США в связи с развитием направления по созданию интеллектуальных обучающих систем (intelligent tutoring system) - обучающих программ, способных не только распознавать ответ, но и оценивать ход мышления обучающегося при выполнении задания [79]. Такие программы отслеживают действия обучающегося и заносят собранные данные в электронный журнал для последующего анализа.

В 2008 году группа ученых организовала первую международную конференцию EDM'08, посвященную анализу образовательных данных в Монреале. Впоследствии эти конференции стали проводиться ежегодно. Новый импульс развитию направления придало появление массовых публичных онлайн-курсов с большой аудиторией и широкими возможностями по сбору данных, таких как Академия Хана, Coursera, edX, Udacity и др. [81].

Развитие методов анализа образовательных данных так или иначе направлено на совершенствование образовательного процесса. Бейкер (Baker) и Яцев (Yacef) определяют четыре цели анализа образовательных данных [81].

1. Прогнозирование поведения учащихся в процессе обучения. Цель достигается путем построения модели, содержащей детальную информацию о познавательных возможностях студента, его знаниях, поведении и мотивации к обучению.
2. Разработка новых моделей и методов представления знаний в предметной области. Используя методы анализа образовательных данных, необходимо выбрать такие формы представления учебного материала, которые соответствуют различным стилям обучения и познавательным способностям, определенным у учащихся.
3. Изучение влияния взаимодействия преподаватель - студент или система обучения - студент. То есть: разработка рекомендаций, которые адаптируются к учащимся, обеспечение обратной связи, показывающей

учителю, как строить изложение материала, изучение влияния помощи образовательной среды в процессе обучения на эти результаты обучения.

4. Развитие знаний о феномене обучения и психологии учащихся.

Специфика анализа образовательных данных заключается в изучаемых данных. Они могут иметь сложную структуру (иерархическую, семантическую), что затрудняет анализ традиционными методами. Это могут быть данные, по которым работники образования обычно не работают.

Таким образом, анализ журналов использования веб-сайта базы данных системы образования может предоставить следующую информацию [82]:

1. количество посещений страниц и количество посетителей (веб- сайт системы электронного обучения);
2. где учащиеся входят и выходят с веб-сайта системы образования;
3. самые посещаемые страницы;
4. браузеры, используемые студентами;
5. частота обращения к определенным страницам во времени (для сайта в целом и для отдельной страницы);
6. географическое расположение доступа пользователей к системе;
7. количество посещений сайта отдельным обучающимся за определенный период времени и их продолжительность;
8. самые популярные ключевые слова для поиска информации в системе;
9. количество просмотров/загрузок обучающих материалов;
10. количество различных страниц, просмотренных студентом в течение сеанса;
11. статистические показатели общения на форуме системы образования (количество сообщений, средний размер записи, популярность поднятой темы, количество запросов по сравнению с другими учащимися с вопросами в отношениях «ученик-учитель»);

12. список электронных ресурсов, используемых, скачиваемых, изученных студентом;

13. количество учебного материала, которое студент читает перед выполнением задания.

Такие данные, в частности, может предоставить система Moodle. В работах [83, 84] описывается опыт сбора данных из этой системы управления обучением.

При анализе образовательных данных широко используются традиционные методы Data mining, основными из которых являются: классификация, кластеризация, поиск связующих правил, поиск последовательных шаблонов, а также поиск текстов (text mining). Указывается, какой метод и в каких задачах используется чаще всего [82]. Все эти методы можно разделить на три большие группы.

Традиционные методы Data mining

Прогнозирования. Целью прогнозирования является создание модели, предполагающей значение интересующей величины (выходной переменной) по значениям, получаемым независимыми переменными (предикторами). Если выходные переменные принимают непрерывные значения, то мы имеем дело с регрессией, ищем отношения между входными и выходными переменными. Если выходная переменная имеет предельное дискретное множество значений (классов), то речь идет о задаче классификации - когда выходная переменная присваивается тому или иному классу в зависимости от значений входных переменных.

Прогноз предполагает, что часть данных задана, то есть соответствующие значения входных переменных и выходных переменных известны для них. Алгоритм регрессии или классификации, «выученный» из этого набора данных, может предсказать выходную переменную для новых (не отмеченных) данных.

Например, исходя из учебных материалов (продолжительность рабочих сессий, количество просмотров и т. д.) и итоговой оценки, полученной студентом на экзамене конкретного преподавателя, можно предсказать оценку, которую студент с подобной учебной активностью получит на экзаменах.

Определение структуры (structure discovery). Алгоритмы определения структуры (structure discovery) пытаются определить структуру в данных, не используя априорные представления об этом. Наиболее популярной группой таких алгоритмов являются кластерные алгоритмы. Кластеризация является логическим продолжением идеи классификации. Его особенность заключается в том, что классы объектов изначально не predetermined, а при классификации некоторые наборы данных должны быть помечены, прежде чем приступить к построению модели. Результатом кластеризации является деление множества объектов на группы близких в определенном смысле объектов.

Определение отношений. Целью определения отношений (relationship mining) является установление отношений между переменными в наборе данных с большим количеством переменных. Например, вы можете попытаться определить, какие переменные зависят от интереса или в какой из пар связанных переменных связь сильнее, чем в других. Определение отношений в анализе образовательных данных часто используется в форме поиска связующих (ассоциативных) правил (association rule mining) или поиска последовательных шаблонов (sequential pattern mining).

В поисках правил связывания целью является нахождение правил типа "if-elif", если (if) некоторые переменные получают некоторый набор значений, то (elif) другая переменная имеет определенное значение. Например, если в транзакции встречается набор товаров А, можно сделать вывод, что в этой транзакции должен появиться набор товаров В.

3.2. Архитектура цифрового сервиса прогнозного анализа качества образования в университете

Необходимость автоматизации информационных процессов связана с увеличением объема информации в информационной системе (ИС) организации, необходимостью ускорения и использования более сложных методов их обработки.

Основные задачи автоматизации информационных процессов [85]:

- снижение трудозатрат при выполнении традиционных процессов и операций;
- удаление ежедневных операций;
- ускорение процессов обработки и преобразования информации;
- расширение возможностей осуществления статистического анализа и повышение точности учетно-отчетной информации;
- повышение оперативности и качественного уровня обслуживания пользователей;
- обновление или полная замена элементов традиционных технологий;
- расширение возможностей организации за счет применения новых информационных технологий и эффективное использование информационных ресурсов организации;
- упрощение возможностей широкого обмена информацией, предоставление услуг, эффективное участие в кооперативных и интеграционных системах.

Одной из целей исследовательской работы является улучшение качества обслуживания посредством разработки чат-бота в мессенджере «Telegram» для автоматизации процесса обработки обращений пользователей сервиса анализа качества образования. Основной программный код чат-бота будем писать на языке программирования Python.

Чат-бот — это программа, которая автоматически отвечает человеку в переписке, реагируя на ключевые слова по заданному сценарию. Это экономит много времени на техническое обслуживание и позволяет сосредоточиться на

важных задачах. Еще одно преимущество бота в том, что он работает круглосуточно [86].



Рис.3.1. Интерфейс Чат-бота

Чат-бот - это умное приложение, которое выполняет различные функции в мессенджерах.

Функции чат-бота [87]:

1. Поддержки пользователей.
2. Клиентский сервис.
3. Маркетинг.
4. Работа внутри компании.
5. Рекрутинг.

Существует несколько вариантов классификации чат- ботов, но, проанализировав их все, можно выделить два типа: бизнес-классификацию приложений чат-ботов и техническую классификацию чат-ботов. Диаграмма бизнес-классификации показана на рис. 3.2.



Рис.3.2. Классификация чат-ботов

Подробнее рассмотрим каждый тип чат-бота:

- Разговорные чат-боты предназначены для общения, очень похожего на общение с обычным человеком, без конкретной цели.
- Чат-боты помощники - необходимая информация получается из ответов пользователей исходя из конкретных целей.
- Чат-боты Q&A - принцип работы: один вопрос – один ответ.

Основные элементы интерфейса чат-ботов универсальны и присущи каждому приложению обмена сообщениями. Как говорится в документации Telegram API, бот может общаться с серверами Telegram двумя способами:

1. `getUpdates-pull`: бот постоянно обращается на сервер Телеграмм, проверяет наличие новых сообщений;
2. `setWebhook-push`: каждый раз, когда приходят новые сообщения, сервер Телеграмм отправляет их в бот.

Разница между этими двумя методами показана на рис. 3.3.

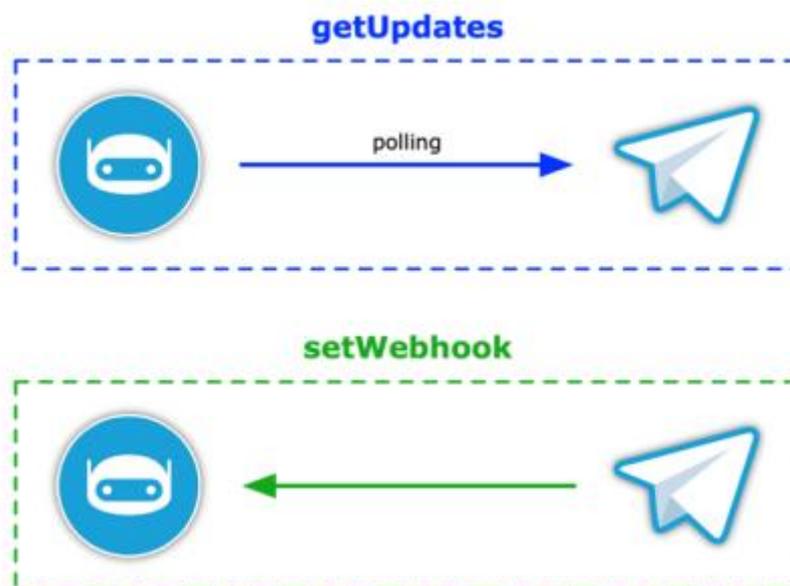


Рис.3.3. Способы взаимодействия с ботом

Очевидно, что второй способ (`setWebhook`) рационален для всех участников процесса. Однако это имеет неоднозначную сложность: кто-то должен получить сообщение от Telegram на стороне бота, то есть веб-сервер или его эквивалент необходимы. Кроме того, сертификат SSL необходим для того, чтобы соединение проходило через защищенный канал HTTPS. Это необходимая и основная часть взаимодействия веб-сервера.

Программное обеспечение включает веб-сервер и Телеграмм-бот.

Основные идеи, реализованные в приложении, приведены ниже:

1. Обработка пользовательских сообщений в режиме реального времени;
2. Оперативное предоставление пользователю возможных вариантов ответа;
3. Обработка пользовательского решения и отправка ответа в режиме реального времени.

Веб-сервер, в свою очередь, состоит из серверного программного обеспечения Телеграм-бота. Т.е. связь с базой данных, выполнение команд, которые запрашивает пользователь.

3.3. Реализация цифрового сервиса прогнозного анализа качества образования в университете на компьютере

Программа- это набор команд и инструкций, написанных на языке, понятном компьютеру. Программа состоит из четких командных схем и инструкций, направленных на выполнение определенной задачи на компьютере. Программирование процесса написания программ, а человек, пишущий программы, называется программистом. Мы говорим, что язык программирования - это язык, который Компьютер может понять. Компьютерная программа это - набор команд с логической структурой и последовательностью для управления компьютером.

Язык программирования- это формальный язык, который связывает человека и компьютер, которые обрабатывают данные по установленным правилам.

Язык программирования это - искусственный язык. Отличается небольшим словарным запасом, строгим соблюдением правил написания кода. Процессор напрямую не принимает программ, написанных на языке программирования. Для этого нужны трансляторы (переводчики), которые переводят программы на язык процессора. Существует 2 вида трансляторов.

1. Компилятор.
2. Интерпретатор.

Для перевода программы, написанной на языке программирования, в машинный код используется компилятор. Интерпретаторы используются для последовательного выполнения программы, переведенной на машинный язык, с поэтапным анализом на процессоре. Оба компилятора и интерпретатора являются программными. Есть разница между компилятором и интерпретатором. Компилятор - полностью переводит программу в машинный код и выполняет ее. Интерпретатор - выполняется с пошаговым анализом, как было сказано выше.

Integrated Development Environment (IDE) (интегрированная среда разработки программ) - комплекс программ для разработки программного обеспечения. Система IDE имеет несколько компонентов (рис. 3.4).

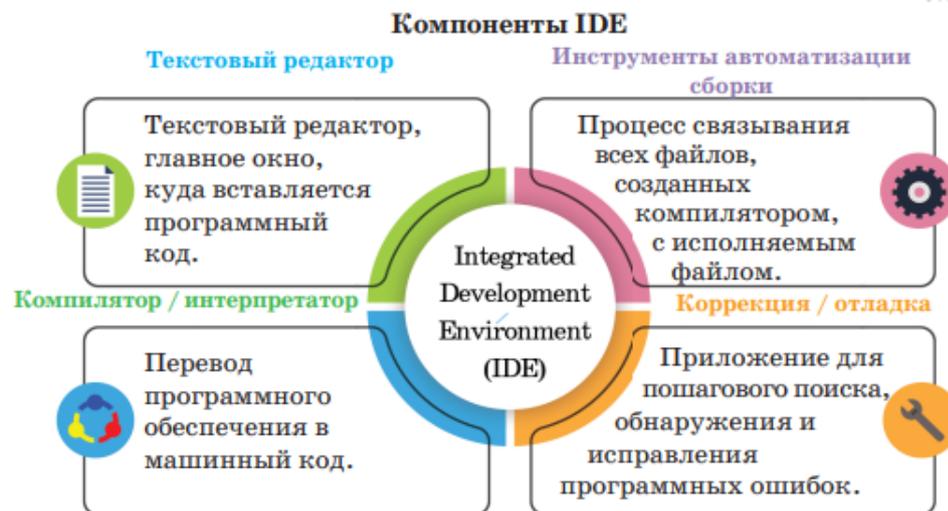


Рис.3.4. Компоненты IDE

Интегрированная среда разработки программ включает такие программы, как C++, Pascal, Java, Python.

При разработке помощника чат-бота сервиса прогнозного анализа качества образования в мессенджере Telegram был выбран язык программирования Python для написания кода программы.

Python - это язык программирования общего назначения высокого уровня. Он также используется для разработки веб-приложений. Язык направлен на повышение производительности разработчиков и читабельности кода.

Python поддерживает различные парадигмы программирования: структурную, объектно-ориентированную, функциональную, императивную и аспектно-ориентированную. Язык включает динамический набор текста, автоматическое управление памятью, самоконтроль, механизм обработки исключений, поддержку многопоточных вычислений и практическую структуру данных высокого уровня [89].

Преимущества Python [90]:

- открытая разработка;
- читать очень легко, особенно на начальном этапе;
- особенности синтаксиса помогают программисту лучше писать;
- читаемый код;
- предлагает инструменты быстрого прототипирования и динамической семантики;
- есть большое сообщество, которое положительно относится к новичкам;
- многие полезные библиотеки и языковые расширения легко использовать в своих проектах благодаря механизму импорта и программным интерфейсам;
- модульные механизмы хорошо продуманы и их легко использовать.

Недостатки Python:

- многопоточная поддержка не очень хорошая;
- В Python не так много качественных программных проектов по сравнению с другими универсальными языками программирования, такими как Java;
- отсутствие коммерческой поддержки инструментов разработки (даже если эта ситуация со временем изменится);
- исходные ограниченные инструменты для работы с базами данных.

Выбор идеального редактора для работы является сложной задачей, которая включает в себя тестирование, личный выбор и окончательное решение. При разработке помощника чат-бота сервиса прогнозного анализа качества образования в мессенджере Telegram был выбран редактор PyCharm для написания кода программы.

PyCharm - это специальная интегрированная среда разработки для Python. Разработан чешской компанией JetBrains. Редактор разработан специально для Python, поэтому он имеет множество функций, таких как автозапуск и проверка

кода, подсветка ошибок, исправления, отладка, система управления версиями и рефакторинг. IDE доступна в Microsoft Windows, Linux и MacOS. Есть бесплатные и платные профессиональные варианты. Профессиональная IDE имеет несколько дополнительных функций, но для большинства задач достаточно бесплатной версии.

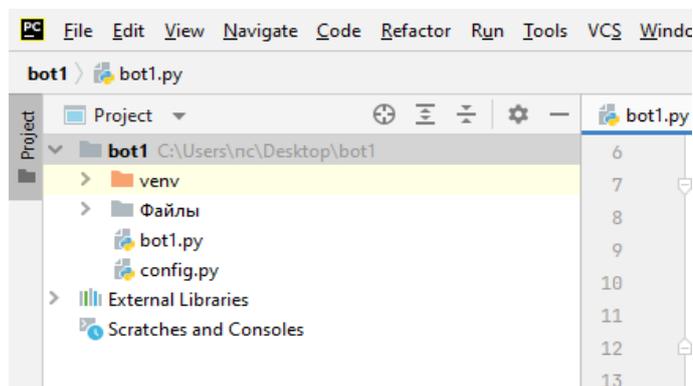


Рис.3.5. Окно редактора PyCharm

Сервис прогнозного анализа качества образования в мессенджере Telegram написан на языке программирования Python, рассмотрим фрагмент программного кода для интерфейса настроек чат-бота, показанный на рисунке ниже.

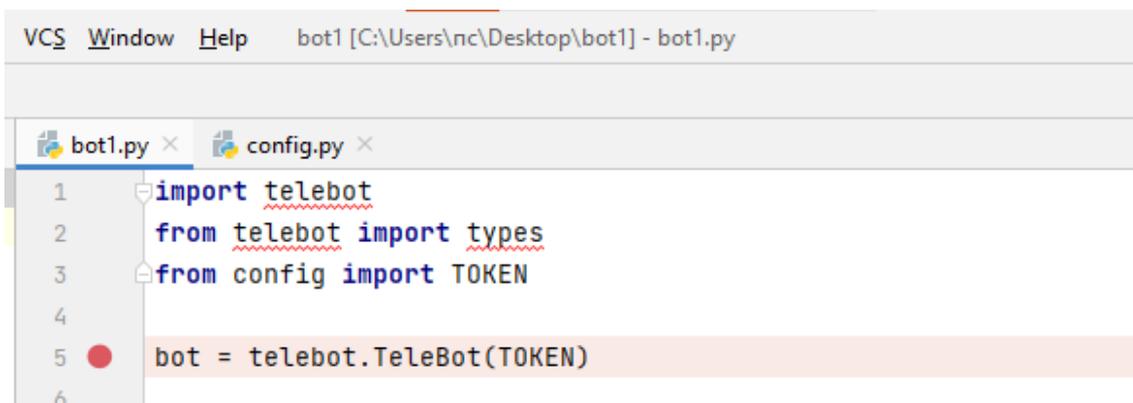


Рис.3.6. Импорт основных библиотек

```

7 m_tizimi = [['Амина Жаулыбаева', 'тел: 8222222222'],
8             ['Ерасыл Сарсенбаев', 'тел: 8777777777'],
9             ['Калабаева Гулмира', 'тел: 8999999999'],
10            ['Манас Ержанов', 'тел: 8333333333'],
11            ['Ринат Адилбеков', 'тел: 8788888888'],
12            ['Серик Нурғалиев', 'тел: 8444444444']]
13 sabaq = ['Алгебра', 'Геометрия', 'Физика', 'Информатика', 'Биология', 'Химия']
14 aty = ['Оразбаев Нұрлан', 'Сабитова Әйгерім', 'Нұрланов Нұрасыл', 'Таңатарова Гүлназ', 'Нұр:
15 name = ['Оразбаев Нурлан', 'Сабитова Айгерим', 'Нурланов Нурасыл', 'Танатарова Гүлназ', 'Н'
16 bagalar = [[5,4,5,5,5],[4,4,4,5,4,4],[5,5,5,5,5],[4,4,4,4,4],[4,5,4,3,4,5],[3,3,3,
17
18 m_id = ['m001', 'm002', 'm003', 'm004', 'm005', 'm006', 'm007', 'm008', 'm009', 'm010']
19
20 d_emes = ['1-а', '1-б', '1-в', '2-а', '2-в', '3-а', '3-б', '3-в', '4-а', '4-б', '4-в',

```

Рис.3.7. Список переменных, используемых в коде программы.

```

22 @bot.message_handler(commands=['start'])
23 def start(message):
24     user = message.from_user.first_name
25     mess = 'қайырлы күн, ' + user + '!' + '\n' + 'Мен университеттің оқу сапасын та
26     mess = mess + 'Добрый день, ' + user + '!' + '\n' + 'Я бот-помощник центра г
27     bot.send_message(message.chat.id, mess)
28
29     id = user + ', ID-кодты енгізіңіз:' + '\n' + '\n' + user + ', введите ID-код'
30     bot.send_message(message.chat.id, id)
31

```

Рис.3.8. Начало работы бота.

```

32 @bot.message_handler(content_types=['audio', 'voice', 'photo', 'video', 'st
33 def mess(message):
34     bot.send_message(message.chat.id, 'Бұл типтегі файлдарға жауап беру қаз
35     + 'В данный момент невозможно ответить на файлы этого
36

```

Рис.3.9. Работа с типами файлов

```

37 @bot.message_handler()
38 def get_user_text(message):
39
40     if message.text in m_id:
41         user = message.from_user.first_name
42         markup = types.ReplyKeyboardMarkup(resize_keyboard=True, row_width=1)
43         kz = types.KeyboardButton('қазақша')
44         ru = types.KeyboardButton('Русский')
45         markup.add(kz, ru)
46         bot.send_message(message.chat.id, user + ', сервиспен жұмысты бастау үшін
47

```

Рис.3.10. Выбор языка интерфейса сервиса.

```
elif message.text == 'Русский' or message.text == 'Главное меню':
    markup = types.ReplyKeyboardMarkup(resize_keyboard=True, row_width=1)
    user = message.from_user.first_name
    t_tizim = types.KeyboardButton('Список групп')
    m_tizim = types.KeyboardButton('Список преподавателей')
    serv = types.KeyboardButton('Сервисы')
    sup = types.KeyboardButton('Написать в службу поддержки')
    lang = types.KeyboardButton('Выбор языка')
    exit = types.KeyboardButton('Выход')
    markup.add(serv, t_tizim, m_tizim, sup, lang, exit)
    bot.send_message(message.chat.id, user + ', пожалуйста выберите пункт
```

Рис.3.11. Создание кнопок

```
elif message.text == 'Оценки':
    for i in range(len(name)):
        esim = name[i]
        esim1 = f'ФИО студента: <b>{esim}</b>'
        b = esim1 + '\n'
        pan = 'Оценки по предметам:'
        b = b + f'<b>{pan}</b>' + '\n'
        for j in range(len(bagalar[i])):
            b = b + sabaq[j] + ' - ' + str(bagalar[i][j]) + '\n'
        bot.send_message(message.chat.id, b, parse_mode='html')

photo = open('Файлы\Оценки студентов.png', 'rb')
bot.send_photo(message.chat.id, photo)

markup = types.InlineKeyboardMarkup()
markup.add(types.InlineKeyboardButton('Оценки студентов', url='http://...'))
bot.send_message(message.chat.id, 'Файл', reply_markup=markup)
```

Рис.3.12. Ответы на запросы

```

elif message.text=='Список групп':
    user = message.from_user.first_name
    markup = types.ReplyKeyboardMarkup(resize_keyboard=True, row_width=3)
    a1 = types.KeyboardButton('1-А')
    b1 = types.KeyboardButton('1-Б')
    v1 = types.KeyboardButton('1-В')
    a2 = types.KeyboardButton('2-А')
    b2 = types.KeyboardButton('2-Б')
    v2 = types.KeyboardButton('2-В')
    a3 = types.KeyboardButton('3-А')
    b3 = types.KeyboardButton('3-Б')
    v3 = types.KeyboardButton('3-В')
    a4 = types.KeyboardButton('4-А')
    b4 = types.KeyboardButton('4-Б')
    v4 = types.KeyboardButton('4-В')
    menu = types.KeyboardButton('Главное меню')
    markup.add(a1, b1, v1, a2, b2, v2, a3, b3, v3, a4, b4, v4, menu)
    bot.send_message(message.chat.id, user + ', выберите нужную группу из списка групп:

```

Рис.3.13. Выбор групп

3.4. Методические указания по использованию цифрового сервиса прогнозного анализа качества образования.

Для запуска сервиса прогнозного анализа качества образования в мессенджере Telegram в поисковую строку браузера нужно написать ссылку <http://t.me/BSTaldauBot> или @BSTaldauBot в строку поиска мессенджера Telegram.

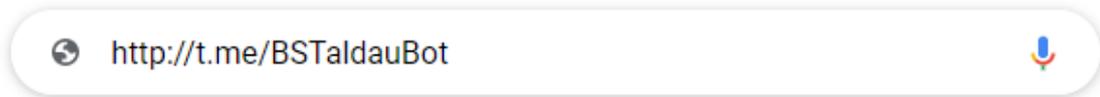


Рис.3.14. Запуск сервиса в браузере



Рис.3.15. Запуск сервиса в мессенджере Telegram

Чтобы запустить сервис необходимо нажать кнопку Старт.

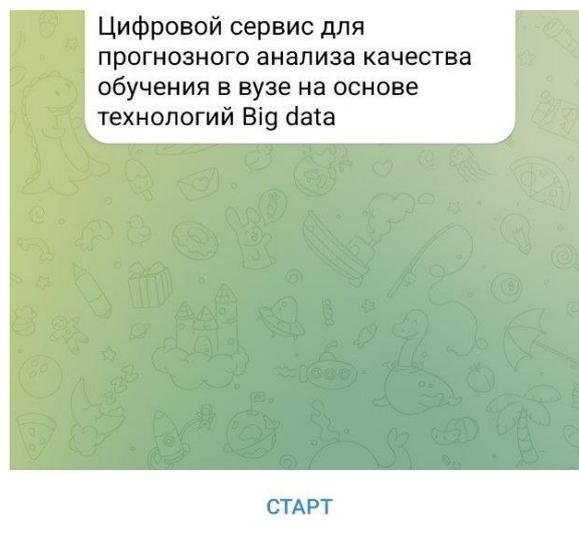


Рис.3.16. Запуск сервиса

После запуска сервиса появится сообщение о вводе ID-кода пользователя. При правильном вводе ID- кода пользователя работа с сервисом будет продолжена.



Рис.3.17. Ввод ID-кода пользователя

После корректного ввода ID-кода пользователя предоставляются языковые настройки для продолжения работы с сервисом.

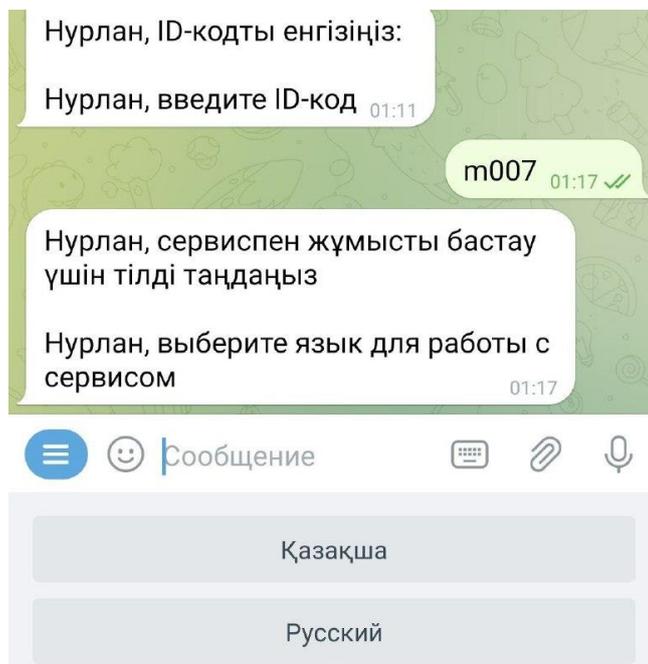


Рис.3.18. Настройки языка

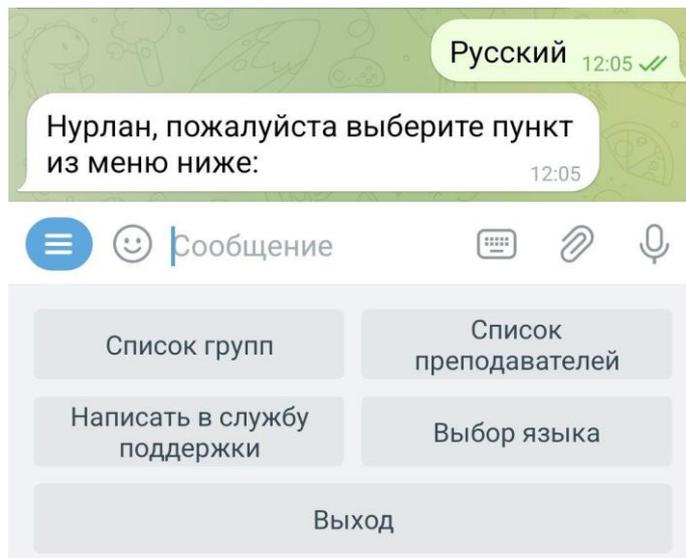


Рис.3.19. Главное меню

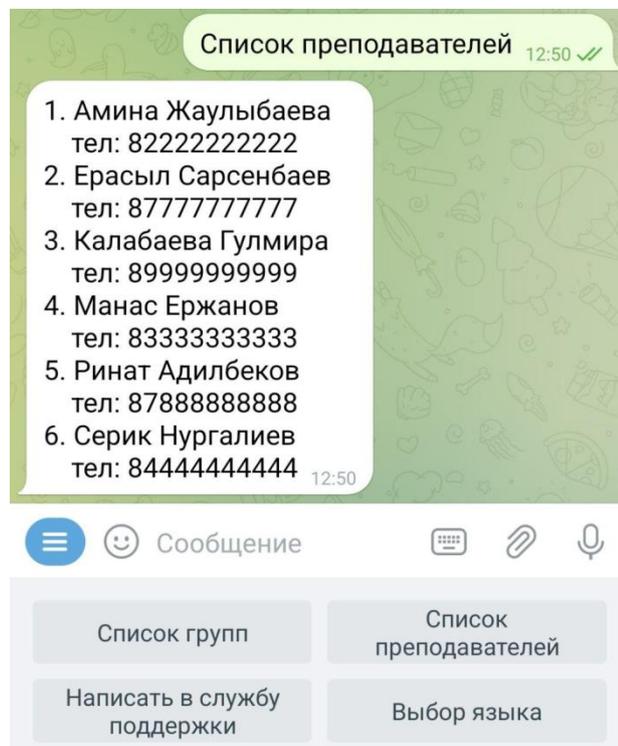


Рис.3.20. Список преподавателей

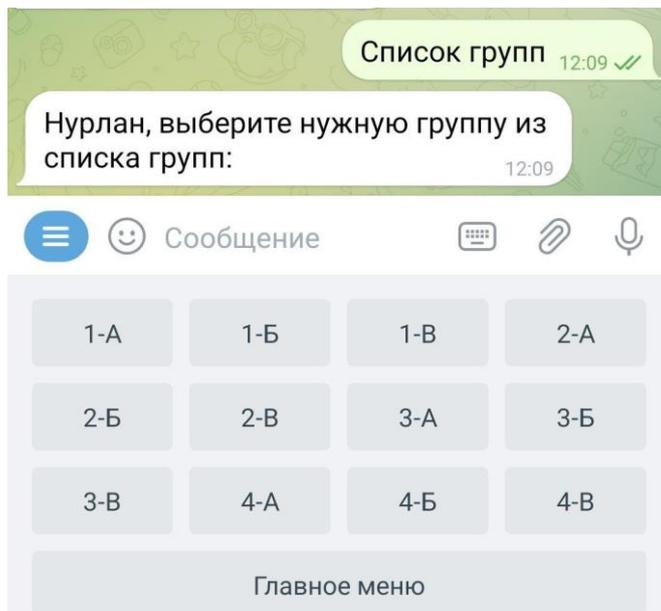


Рис.3.21. Список групп

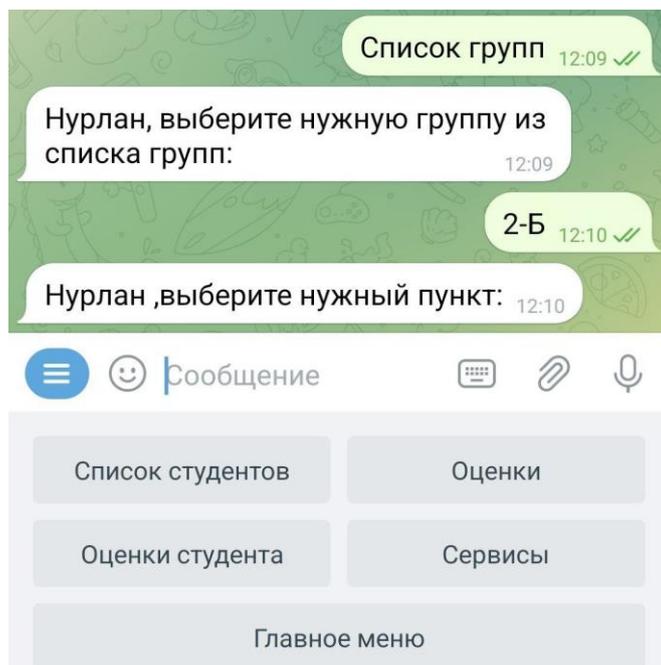


Рис.3.22. Выбор группы

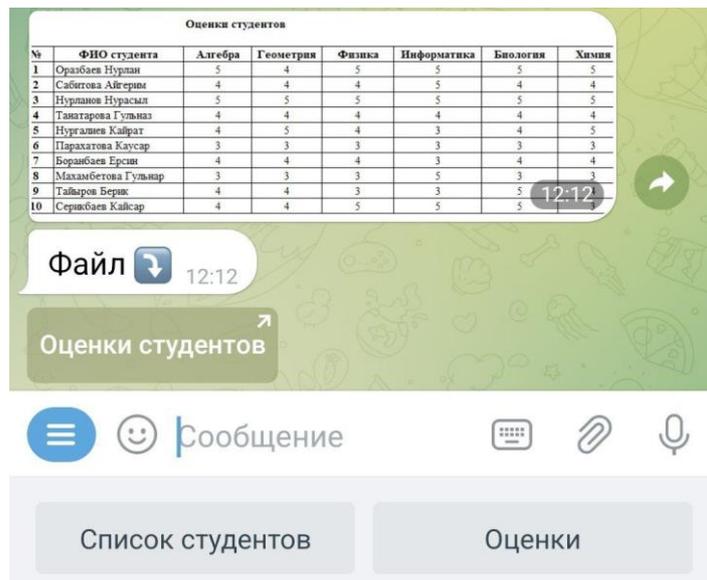


Рис.3.23. Оценки студентов группы

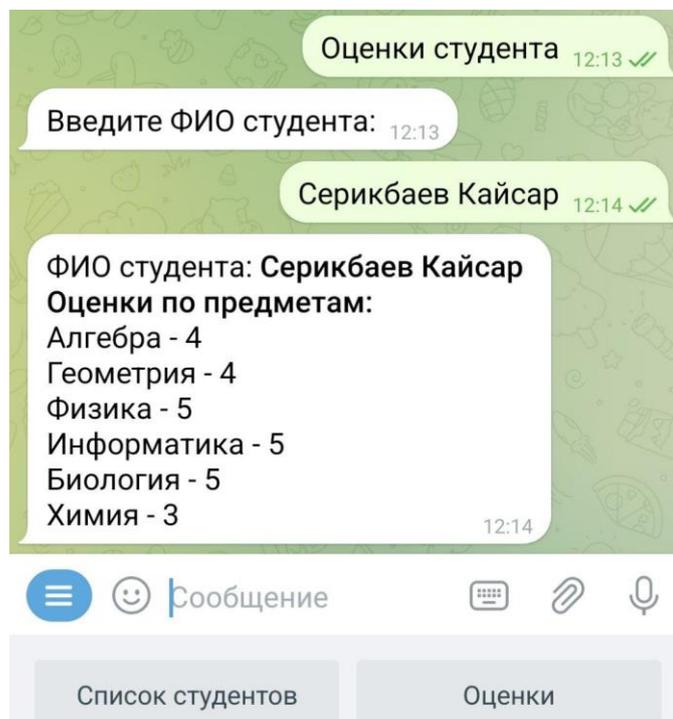


Рис.3.24. Оценки студента

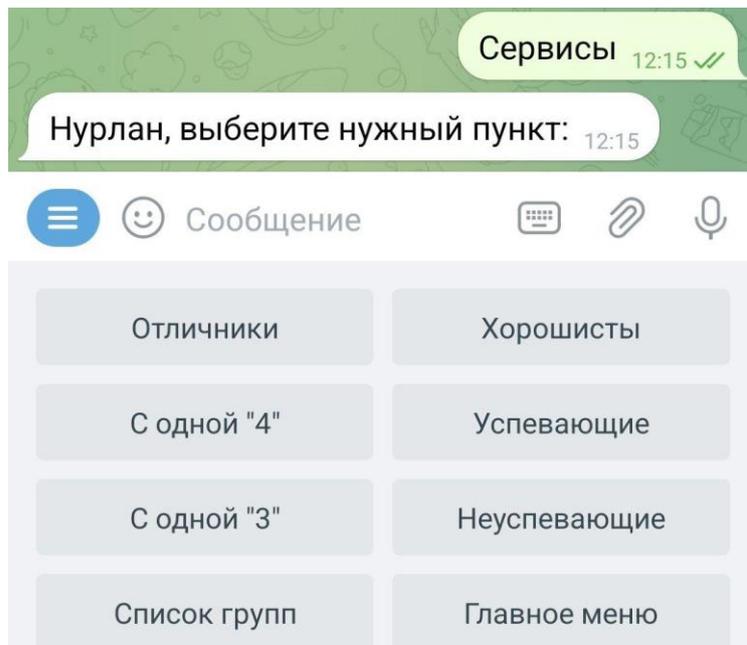


Рис.3.25. Сервисы

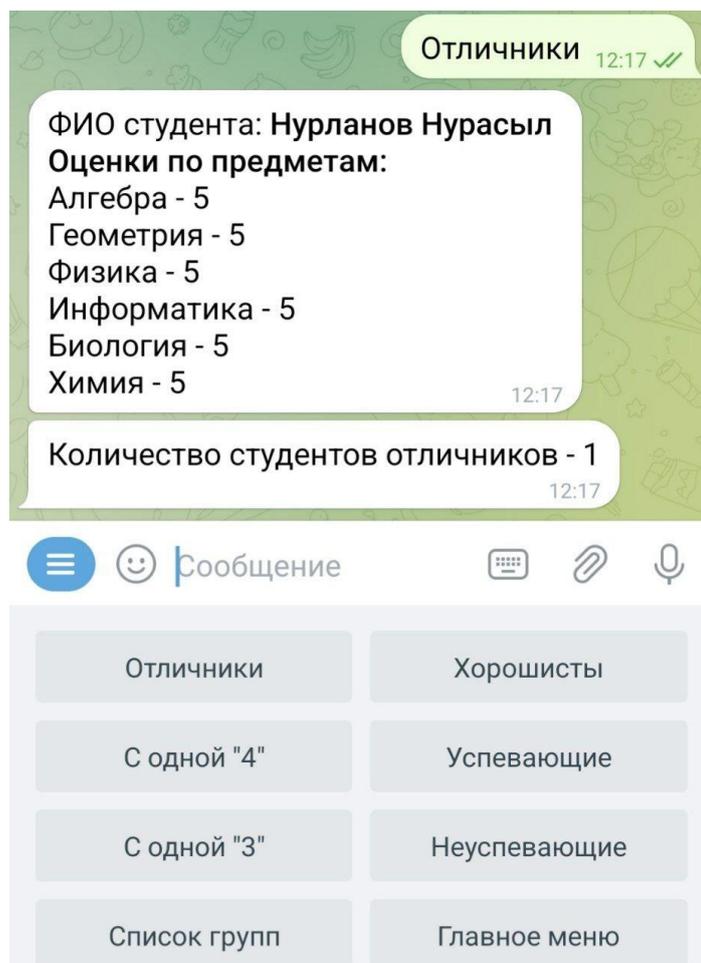


Рис.3.26. Отличники

Заключение

Big Data открывает новые горизонты в современном образовании, с развитием этих технологий применение цифрового сервиса прогнозного анализа качества образования в образовании выходит на высокий уровень, позволяющий выделить студентов группы риска или студентов, достойных особых достижений. Это позволит отслеживать подобные ситуации и помогать им успешно продвигаться по индивидуальной образовательной траектории, а также избежать ситуации потери места в университете. Анализ данных о качестве обучения может направить студентов на выбор образования и карьеры, которые соответствуют их личностным качествам и интересам в будущей перспективе.

Прогностический анализ качества образования на основе технологий Big Data может стать мощным инструментом трансформации обучения, переосмысления подходов, сокращения имеющихся пробелов и адаптации опыта для повышения эффективности системы образования.

Подводя итог, можно сказать, что использование больших данных во всех областях позволяет управлять и использовать множество функций. Анализ данных позволяет работать с индивидуальными программами студентов, персонализировать обучение. Данные показывают, с каким содержанием студент

взаимодействует, как происходит это взаимодействие, где он заинтересован и где ему скучно, с кем и как он взаимодействует в процессе обучения, как прохождение определенного курса повлияло на результаты обучения, на каком этапе обучения ему нужна помощь. При таком подходе обучение ориентировано на личность и целенаправленно направлено на адаптационный подход. Также на основе больших данных образовательная аналитика меняет представление о формате образовательных программ. Тексты, используемые в образовательном процессе, могут быть не только оцифрованы, но и переведены на цифровые данные. Пользователи продвигаются вперед с большей свободой в материале, а затем анализируется, как пользователи взаимодействуют с материалом. Результатом такой аналитики является изменение содержания. Поэтому образовательная программа превращается из утвержденного текстового формата в формат определенного набора онлайн-контента, который динамически изменяется путем анализа данных, возникающих в результате взаимодействия с онлайн-контентом учащихся. Появится умная программа, умная учебная программа. Можно предположить, что программы учебных курсов также претерпят изменения, а также может измениться образовательный процесс, образовательные результаты, подходы к мониторингу и оценке. Мониторинг будет регулярным. Интерес студентов к постоянному мониторингу связан с тем, что анализ данных позволяет сделать его учебный план индивидуальным, интерес учителей связан с возможностью получения информации о продуктивных группах, обратной связью с контентом, созданным студентами, эффективным распределением ресурсов для преподавателей. Оценка образовательных результатов может быть объединена независимо или коллективно на основе всех данных учащихся, полученных из всего взаимодействия. Оценка проводится для грамотного расширения образовательной программы студента. Постоянно фиксируется динамика образовательных результатов, на основе этих данных формируются закономерности, по которым можно судить о развитии студента.

Таким образом основные результаты исследования заключаются в следующем:

- Проведенный анализ возможностей новых информационных технологий в использовании технологий больших данных в отечественных и зарубежных исследованиях показал перспективность их применения в образовании;

- На основе технологий Big Data разработана модель цифрового сервиса прогнозного анализа качества образования в университете, включающая в себя структуру, компоненты, функции и результаты;

- Разработан цифровой сервис предиктивного анализа качества обучения учащихся на основе технологии Big Data, который может быть внедрен в реальную образовательную практику.

Библиографический список.

1. Ammar Almasri, Erbug Celebi, Rami S. Alkhaldeh EMT: Ensemble MetaBased Tree Model for Predicting Student Performance. 2019. 11-б.

2. Быстрова Т. Ю., Ларионова В. А., Сеницын Е. В., Толмачев А. В. Учебная аналитика MOOK как инструмент прогнозирования успешности обучающихся // Вопросы образования. – 2018. – № 4. – С. 139–166.
3. Вилкова К. А., Захарова У. С. Учебная аналитика в традиционном образовании: её роль и результаты // Университетское управление: практика и анализ. – 2020. – № 24 (3). – С. 59–76
4. Lubna Mahmoud Abu Zohair Prediction of Student's performance by modelling small dataset size // International Journal of Educational Technology in Higher Education. 2019. 16-б.
5. Lynch C. Big data: how do your data grow? // Nature. 2008. Vol. 455. № 7209. P. 28—29.
6. Бебенина Е. В., Елкин О. М. Повышение качества управления образованием с использованием технологии обработки больших данных // Отечественная и зарубежная педагогика. 2020. № 6 (72). С. 22—29
7. Liebowitz J. Thoughts on recent trends and future research perspectives in big data and analytics in higher education // Big data and learning analytics in higher education: Current theory and practice. January, 2016. P. 7—17
8. Аналитика больших данных и machine learning в образовании: 5 кейсов из вузов. URL: <https://www.bigdataschool.ru/blog/big-data-analytics-education-cases.html>. (Дата обращения: 02.03.2022)
9. Прогнозировать и предотвращать отчисления [https:// skillbox. ru/ media/ education/ bolshie- dannye-v- vysshem- obrazovaniy/](https://skillbox.ru/media/education/bolshie-dannye-v-vysshem-obrazovanii/) (Дата обращения: 07.03.2022)
10. Е. Ю. Огурцова, Р. Н. Фадеев. Большие данные и цифровая аналитика в университетском образовании. [https:// cyberleninka. ru/ article/n/ bolshie- dannye-i- tsifrovaya- analitika-v- universitetskom- obrazovanii](https://cyberleninka.ru/article/n/bolshie-dannye-i-tsifrovaya-analitika-v-universitetskom-obrazovanii) (Дата обращения: 07.03.2022)
11. Liebowitz J. Thoughts on recent trends and future research perspectives in big data and analytics in higher education // Big data and learning analytics in higher

- education: Current theory and practice. – January, 2016. – P. 7– 17. doi: 10.1007/978-3-319-06520-5_2
12. Zawacki-Richter O., Latchem C. Exploring four decades of research in computers & education // *Computers and Education*. – 2018. – № 122. – P. 136–152. doi: 10.1016/j.compedu.2018.04.001.
 13. Vieira C., Parsons P., Byrd V. Visual learning analytics of educational data: A systematic literature review and research agenda // *Computers and Education*. – 2018. – № 122. – P. 119–135. doi: 10.1016/j.compedu.2018.03.018
 14. Buniyamin N., Mat U. B., Arshad P. M. Educational data mining for prediction and classification of engineering students achievement // Paper presented at the 2015 IEEE 7th International Conference on Engineering Education, ICEED 2015. – 2018. – P. 49–53. doi: 10.1109/ICEED.2015.7451491
 15. De Almeida Neto F. A., Castro A. A reference architecture for educational data mining // Paper presented at the Proceedings Frontiers in Education Conference, FIE. – 2017. – October. – P. 1–8. doi: 10.1109/FIE.2017.8190728.
 16. Buenaño-Fernández D. B., Luján-Mora S. Comparison of applications for educational data mining in engineering education // Paper presented at the EDUNINE 2017 – IEEE World Engineering Education Conference: Engineering Education – Balancing Generalist and Specialist Formation in Technological Carriers: A Current Challenge, Proceedings. – 2017. – P. 81–85. doi: 10.1109/EDUNINE.2017.7918187.
 17. Yu X., Wu S. Typical Applications of Big Data in Education // *International Conference of Educational Innovation through Technology (EITT)*. – Wuhan, 2015. – P. 103–106. doi: 10.1109/EITT.2015.29.
 18. Farokhmehr M., Fatemi S. O. Implementing machine learning on a big data engine for e-learning // Paper presented at the Proceedings of the European Conference on e-Learning, ECEL. – 2016. – January. – P. 188–193.
 19. Nasiri M., Minaei B., Vafaei F. Predicting GPA and academic dismissal in LMS using educational data mining: A case mining // Paper presented at the 3rd

- International Conference on eLearning and eTeaching, ICeLeT. – 2012. – P. 53–58. doi: 10.1109/ICELET.2012.6333365
20. Moscoso-Zea O., Vizcaino M., Luján-Mora S. Evaluation of methods and algorithms of educational data mining // Paper presented at the 2017 Research in Engineering Education Symposium, REES 2017.
21. Утёмов В. В., Горев П. М. Развитие образовательных систем на основе технологии Big Data // Научно-методический электронный журнал «Концепт». – 2018. – № 6 (июнь). – С. 449–461. – URL: <http://e-koncept.ru/2018/181039.htm>
22. Asha T, Shrvanthi U.M, Nagashree N, Monika M, Building Machine Learning Algorithms on Hadoop for Bigdata // International Journal of Engineering and Technology. 2013. Vol. 3. No. 2. P. 143–147.
23. Qi Zhang, Lu Cheng, and Raouf Boutaba. Cloud computing: stateof-the-art and research challenges. Journal of Internet Services and Applications. 1(1): 2010. P. 7–18
24. URL: www.storm.apache.org, URL: www.spark.apache.org.
25. J. Schmidhuber. Deep learning in neural networks: An overview. Neural Networks, 61: 2015. P. 85–117.
26. Козлова И. В., Саидахмедова М. Б. Аналитика в теоретическом спектре цифровизации высших учебных заведений на основе обзора зарубежных источников // Международный научно-исследовательский журнал. – 2021. – № 4 (106). – С. 123–126.
27. Котова Е. Е. Прогнозирование успешности обучения в интегрированной образовательной среде с применением инструментов онлайн аналитики // Компьютерные инструменты в образовании. – 2019. – № 4. – С. 55–80.
28. Озерова Г. П., Павленко Г. Ф. Прогнозирование успешности студентов при смешанном обучении с использованием данных учебной аналитики // Science for Education Today. – 2019. – Т. 9, № 6. – С. 73–87.

29. Патаракин Е. Д. Совместная сетевая деятельность и поддерживающая её учебная аналитика // Высшее образование в России. – 2015. – № 5. – С. 145–154.
- 30.11. Патаракин Е. Д. Учебная аналитика совместной сетевой деятельности // Школьные технологии. – 2015. – № 4. – С. 80–86.
31. Anthony Dalton, Justin Beer, Sriharshasai Kommanapalli, James S. Lanich. Machine Learning to Predict College Course Success // SMU Data Science Review. – 2018. – Vol. 1. – № 2. – Art. 1
32. Usman Ashfaq, Booma P. M., Raheem Mafas. Managing Student Performance: A Predictive Analytics using Imbalanced Data // International Journal of Recent Technology and Engineering (IJRTE). – 2020. – P. 2277–2283.
33. Vaibhav Kumar M. L. Garg. Predictive Analytics: A Review of Trends and Techniques // International Journal of Computer Applications. – 2018. – P. 31–37.
34. D. Revina Rebecca, Mahammad Ashraf, Partha Pratim Sarkar. A Classification based predictive model to predict students aspiring for higher education // Proceedings of the 2nd National Conference on Advanced Computing Technologies and Applications (NCACTA 2019). – 2019.
35. Vaibhav Kumar M. L. Garg. Predictive Analytics: A Review of Trends and Techniques // International Journal of Computer Applications. – 2018. – P. 31–37.
36. Systems Engineering Thinking Wiki [Электронный ресурс] // Сайт о системной инженерии, проектном управлении и смежных областях. – URL: http://sewiki.ru/Data_Mining (дата обращения: 14.12.2021).
37. Чубукова И. А. Data Mining. – М., Саратов: Национальный Открытый Университет «ИНТУИТ», Ай Пи Ар Медиа, 2020. – 469 с.
38. Федин Ф. О., Федин Ф. Ф. Анализ данных. Инструменты Data Mining. – М.: Московский городской педагогический университет, 2012. – 308 с.
39. Лимановская О. В., Алферьева Т. И. Основы машинного обучения. – Екатеринбург: Издво Урал. ун-та, 2020. – 88 с.

40. Tech Funnel [Электронный ресурс]. – URL: <https://www.techfunnel.com/hr-tech/top-3-examples-of-predictive-analytics-in-hr/> (дата обращения: 16.12.2021)
41. SmartDataCollective [Электронный ресурс] // Сайт о больших данных, аналитике, искусственном интеллекте и облаке. – URL: <https://www.smartdatacollective.com/amazonusing-big-data-analytics-read-your-mind/> (дата обращения: 18.12.2021).
42. Железнов М. М. Методы и технологии обработки больших данных. – М.: Издательство МИСИ – МГСУ, 2020. – 46 с
43. Прошкина, Е. Н. Анализ и прогнозирование успеваемости студентов на основе радиальной базисной нейронной сети / Е. Н. Прошкина, И. Ю. Балашова. — Текст : непосредственный // Технические науки: традиции и инновации : материалы III Междунар. науч. конф. (г. Казань, март 2018 г.). — Казань : Молодой ученый, 2018. — С. 24-28. — URL: <https://moluch.ru/conf/tech/archive/287/13683/> (дата обращения: 05.01.2022).
44. Певченко, С. С. Сравнительный анализ алгоритмов нейронной сети и деревьев принятия решений модели интеллектуального анализа данных / С. С. Певченко, В. А. Блужин. — Текст : непосредственный // Молодой ученый. — 2016. — № 28 (132). — С. 148-154. — URL: <https://moluch.ru/archive/132/36999/> (дата обращения: 07.01.2022).
45. В.А. Шевченко Прогнозирование успеваемости студентов на основе методов кластерного анализа // Вестник ХНАДУ. 2015. №68. URL: <https://cyberleninka.ru/article/n/prognozirovanieuspevaemosti-studentov-na-osnove-metodov-klasternogo-analiza> (дата обращения: 08.01.2022).
46. ПРОГНОЗИРОВАНИЕ ПЕРСОНАЛЬНОЙ УСПЕВАЕМОСТИ СТУДЕНТОВ В ВУЗЕ. Будаева А.А. В сборнике: ИТ-Технологии: развитие и приложения XV Ежегодная Международная научнотехническая конференция: Сборник докладов. 2018. С. 9-16.

47. Al-Shehri H. et al. Student performance prediction using support vector machine and k-nearest neighbor // 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE). – IEEE, 2017. – С. 1-4
48. Ясинский И.Ф. Опыт прогнозирования успеваемости студентов при помощи нейросетевой технологии // Вестник ИГЭУ. – 2007. – № 4. – С. 1 – 4.
49. Прошкина Е. Н., Балашова И. Ю. Анализ и прогнозирование успеваемости студентов на основе радиальной базисной нейронной сети [Текст] // Технические науки: традиции и инновации: материалы III Междунар. науч. конф. (г. Казань, март 2018 г.). — Казань: Молодой ученый, 2018. — С. 24-28. — URL <https://moluch.ru/conf/tech/archive/287/13683/> (дата обращения: 22.02.2022).
50. Харламова И.Ю. Прогнозирование успеваемости студентов первого курса по результатам сдачи единого государственного экзамена // Базис. – 2017. – № 1(1). – С. 57 – 59.
51. Губин Е.И. Методика подготовки больших данных для прогнозного анализа // Наука и бизнес: Пути развития, № 3(105) 2020, с. 27-31
52. Breazley, D. Python Cookbook, Third Edition / D. Breasley, В. К. Jones. – USA: O’Reilly Media, 2013. – 688 p.
53. Бережная И. Ф. Проектирование индивидуальной траектории профессионального развития студентов в вузе на основе объектно-субъектного преобразования / И. Ф. Бережная // Вестник ВГУ. Сер.: Проблемы высшего образования. – 2015. – № 2. – С. 89–94.
54. Зеер Э. Ф. Теоретико-прикладные основания прогнозирования профессионального будущего человека / Э. Ф. Зеер, Э. Э. Сыманюк // Фундаментальные исследования. – 2014. – № 9–8. – С. 1863–1869. – Режим доступа: <http://fundamental-research.ru/ru/article/view?id=35157>
55. Зеер Э. Ф. Психологические основы формирования развивающего профессионально-образовательного пространства студентов колледжа / Э.

- Ф. Зеер, И. В. Мешкова, Л. П. Панина. – Екатеринбург : ГОУ ВПО Рос. гос. проф.-пед. ун-т, 2007. – 124 с
- 56.Александрова Е. А. Педагогическое сопровождение старшеклассников в процессе разработки и реализации ИОТ: автореферат диссертации ... доктора педагогических наук / Е. А. Александрова. Тюмень, 2006. 40 с
- 57.Маскаева А. М. Проектирование ИОТ обучающихся / А. М. Маскаева // Инициативы XXI века. 2010. № 3. С. 23–24.
- 58.Суртаева Н. Н. Педагогические технологии в реализации гуманистической концепции образования / Н. Н. Суртаева // Химия в школе. 1997. № 7. С. 13–23.
- 59.Хуторской А. В. Методика личностно-ориентированного обучения. Как обучать всех по-разному?: пособие для учителя / А. В. Хуторской. Москва: Владос-Пресс, 2005. 383 с.
- 60.Якиманская И. С. Требования к учебным программам, ориентированным на личностное развитие школьников / И. С. Якиманская // Вопросы психологии. 1994. № 2. С. 64–67.
- 61.Осипьянц, Т. С. Проектирование индивидуальной образовательной траектории студента среднего профессионального образовательного учреждения как необходимое условие личностно-профессионального становления / Т. С. Осипьянц. — Текст : непосредственный // Актуальные вопросы современной педагогики : материалы X Междунар. науч. конф. (г. Самара, март 2017 г.). — Самара : ООО "Издательство АСГАРД", 2017. — С. 138-142. — URL: <https://moluch.ru/conf/ped/archive/212/11996/> (дата обращения: 08.02.2022).
- 62.Ковалёва Т. М., Кобыща Е. И., Попова (Смолик) С. Ю., Теров А. А., Чередилина М. Ю. Профессия «тьютор». М. ; Тверь : СФК-офис, 2012
- 63.Katalin Fehér Digital identity and the online self: Footprint strategies – An exploratory and comparative research study, Journal of Information Science, 2019

64. Galimova E.G., Konysheva A.V., Kalugina O.A., Sizova Z.M. Digital educational footprint as a way to evaluate the results of students learning and cognitive activity in the process of teaching mathematics, *EURASIA Journal of Mathematics, Science and Technology Education*, 2019, Vol. 15(8).
65. Azcona, D., Hsiao, I., Smeaton, A.F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints, In *User Modeling and User-Adapted Interaction*, 29, 759–788
66. Suhonen S. “Learning analytics: Combining Moodle, Whatsapp and self-evaluation data for better understanding”, in: *Proceedings of the 6th European Conference on Social Media*, eds. W. Popma, S. Francis. UK, Brighton, Academic Publishing, ECSM, 2019, pp. 410–413.
67. Ng E.Y.L., Law N., Yuen A.H.K. “Understanding learner lives through digital footprints”, in: *Proceedings of the Technology, Mind, and Society conference (TECHMINDSOCIETY18)*, ed. A.L. Story. New York, ACM, 2018, Article 26.
68. Мантуленко В.В. Перспективы использования цифрового следа в высшем образовании // Преподаватель XXI век. 2020. №3-1. URL: <https://cyberleninka.ru/article/n/perspektivyispolzovaniya-tsifrovogo-sleda-v-vysshem-obrazovanii> (дата обращения: 20.02.2022).
69. Курбацкий В.Н. Цифровой след в образовательном пространстве как основа трансформации современного университета. URL: <http://elib.bsu.by/handle/123456789/239268> (дата обращения: 21.02.2022)
70. Захарова И.Г., Карпов М.Г., Лобунцов Д.С. Информационно-аналитическая поддержка управления образовательным процессом с использованием данных цифрового следа студента. URL: <https://www.elibrary.ru/item.asp?id=44034512> (дата обращения: 21.02.2022)
71. Gamacho, M., Minelli, J., Grosbeck, G. (2012). Self and Identity: Raising Undergraduate Students' Awareness on Their Digital Footprints, In *Procedia – Social and Behavioral Sciences*, 46, 3176–3181.
72. Хореева Н.К., Дараган С.В. Мониторинговые исследования по повышению качества обучения в вузе. URL: <https://i->

- exam.ru/sites/default/files/Savchenkov_V_Information_technologies.pdf (дата обращения: 23.02.2022)
73. Corbett, A. T. and Anderson, J. R. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* 4, 4, 253—278.
74. Pavlik, P. I., Cen, H. & Koedinger, K. (2009) Performance Factors Analysis — A New Alternative to Knowledge. *Proceedings of the 14th International Conference on Artificial Intelligence in Education, Brighton, UK*, 531—538.
75. Schneider D. et al. Requirements for learning scenario and learning process analytics. 2012. Vol. 2012. P. 1632–1641.
76. Слєпкань З. І. Наукові засади педагогічного процесу у вищій школі: навч. посіб. Київ: Вища школа, 2005. 239 с.
77. Baker R., Siemens G. Educational Data Mining and Learning Analytics. URL: <http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf> (дата обращения: 24.02.2022)
78. Siemens G., Baker R. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration // *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 2012. URL: <http://www.columbia.edu/~rsb2162/LAKs%20reformatting%20v2.pdf> (дата обращения: 24.02.2022)
79. Intellectual'noe upravlenie protsessom obucheniya [Intelligent control of the learning process]. URL: <https://habrahabr.ru/post/194240/> (дата обращения: 24.02.2022)
80. Sitikhadijah M., Zaidatun T. Educational Data Mining: A Review. URL: <https://www.researchgate.net/publication/275542685> (дата обращения: 25.02.2022)
81. Baker R., Yacef K. The state of educational data mining in 2009: A review and future visions // *Journal of Educational Data Mining*. 2009. V. 1, N 1. P. 3-17.

82. Romero C., Ventura S. Educational Data Mining: A Review of the State of the Art // IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews). 2010, V. 40, № 6. P. 601-618. doi: 10.1109/TSMCC.2010.2053532.
83. Romero C., Ventura S., Garcia E. Data mining in course management systems: Moodle case study and tutorial // Computers & Education. 2007. doi:10.1016/j.compedu.2007.05.016.
84. AlAjmi M.F., Khan Sh., Zamani A.S. Using Instructive Data Mining Methods to Revise the Impact of Virtual Classroom in E-Learning // International Journal of Advanced Science and Technology. 2012. V. 45, N 8. P. 125-134.
85. Автоматизированные информационные системы [Электронный ресурс]. URL: http://www.e-biblio.ru/book/bib/01_informatika/Avtomat_info_sistem/sg.html (Дата обращения: 11.03.2022).
86. Чат-бот для Телеграм [Электронный ресурс]. – URL: <https://www.jivo.ru/blog/tutorials-jivo/chat-bot-dlya-telegram.html> (Дата обращения: 18.03.2022).
87. Как использовать чат-боты в бизнесе: 5 идей и 5 кейсов [Электронный ресурс]. – URL: <https://vc.ru/services/93850-kak-ispolzovat-chat-boty-vbiznese-5-idey-i-5-keysov> (Дата обращения: 18.03.2022).
88. Қадырқұлов Р.А., Нұрмұханбетова Г. К. Информатика. Жалпы білім беретін мектептің 6-сынып оқушыларына арналған оқулық / Р.А. Қадырқұлов, Г.К. Нұрмұханбетова. – Алматы: «Алматыкітап баспасы», 2020. 73-бет.
89. Язык программирования Python [Электронный ресурс]. – URL: <https://web-creator.ru/articles/python> (Дата обращения: 19.03.2022).
90. Интернет технологии [Электронный ресурс]. – URL: <https://www.internet-technologies.ru/articles/php-ruby-python> (Дата обращения: 25.03.2022).

- 91.Изимбетов Н.О. VI Международной научно-практической конференции «Наука и образование-2022», Нур-Султан – 2022. С. 246-248
- 92.Изимбетов Н.О. Жоғары оқу орындарында цифрлық ізді пайдалану мүмкіндіктері / Н. О. Изимбетов. — Текст : непосредственный // Молодой ученый. — 2022. — № 16 (411). — С. 519-523. — URL: <https://moluch.ru/archive/411/90607/> (дата обращения: 31.05.2022).