

Санкт-Петербургский государственный университет

КРАГЕЛЬ Полина Игоревна

Выпускная квалификационная работа

Анализ цифрового следа при онлайн-обучении

Уровень образования: бакалавриат

Направление *02.03.03 «Математическое обеспечение и администрирование
информационных систем»*

Основная образовательная программа *СВ.5006.2018 «Математическое обеспечение и
администрирование информационных систем»*

Профиль *Информационные системы и базы данных*

Научный руководитель:
доцент кафедры информационно-аналитических систем, к.ф.-м.н., Е.Г. Михайлова

Рецензент:
доцент Высшей школы цифровой культуры университета ИТМО, Д.Г. Волчек

Санкт-Петербург
2022

Saint Petersburg State University

Polina Kragel

Bachelor's Thesis

Digital footprint analysis in online learning

Education level: bachelor

Speciality *02.03.03 "Software and Administration of Information Systems"*

Programme *CB.5006.2018 "Software and Administration of Information Systems"*

Profile: *Information Systems and Databases*

Scientific supervisor:
C.Sc., docent E.G. Mikhailova

Reviewer:
docent D.G. Volchek

Saint Petersburg
2022

Оглавление

1. Введение	4
2. Постановка задачи	6
3. Обзор предметной области	7
3.1. Анализ образовательных данных	7
3.2. Методы EDM	9
4. Данные	11
4.1. Группы событий	12
4.2. База данных	16
5. Решение	18
5.1. Архитектура решения	18
5.2. Обработка логов	19
5.3. Анализ онлайн-курса	19
5.4. Кластеризация	21
6. Заключение	29
Список литературы	30

1. Введение

В современном обществе все большее значение приобретают цифровые технологии, а их влияние на сферы жизни человека только увеличивается. Сфера образования не является исключением. Использование цифровых технологий в области образования предоставляет множество возможностей, применимых для улучшения учебного процесса, поэтому в нынешних реалиях всё большую популярность приобретает онлайн-образование. В последнее время появилось большое число образовательных платформ, содержащих огромное количество разнообразных онлайн-курсов. В частности, одними из самых популярных онлайн-платформ в России являются «Stepik» и «Открытое образование».

Широкое распространение онлайн-обучения привело к появлению большого количества информации, которое может быть использовано для повышения эффективности учебного процесса и улучшения качества образования [16, 14, 3, 6]. К такой информации может относиться время и порядок обращения учащегося к образовательному контенту, частота возвращения учащегося к образовательным материалам, результаты выполнения тестовых заданий, переходы на страницах и другие данные учебной аналитики. Все это входит в понятие цифровой след.

Анализирование цифрового следа может помочь в улучшении учебного плана онлайн-курса и оптимизации учебного процесса [3, 10, 5]. Также анализ цифрового следа может выявить определенные шаблоны поведения учащегося или определить студентов, подозреваемых в недобросовестном прохождении курса [7, 16]. К тому же, исследование полученных данных позволяет скорректировать курс, выявляя излишне простые или сложные задания, а иногда может помочь выявлять задания, составленные с ошибками.

В связи с широкой применимостью анализа образовательных данных в разных задачах данное направление начало активно развиваться. Анализ цифрового следа в онлайн-обучении на данный момент является одной из самых исследуемых областей. Активное применение

информационных технологий в области образования и бурный рост количества онлайн-курсов этому только способствовали. Соответственно, возникает необходимость в инструментах для анализа цифрового следа обучающегося.

2. Постановка задачи

Целью данной работы является анализ цифрового следа для оптимизации и оценки результатов онлайн-обучения.

Для реализации данной цели необходимо выполнить следующие задачи:

- Изучение предметной области, а именно Educational Data Mining.
- Рассмотрение существующих подходов в EDM.
- Конкретизация значимой информации в логах.
- Проектирование схемы базы данных.
- Обработка логов и их загрузка в базу данных.
- Разработка инструментов для анализа онлайн-курса и определения шаблонов поведения учащихся.

3. Обзор предметной области

3.1. Анализ образовательных данных

Анализ образовательных данных (Educational Data Mining) — это область исследования, в которой применяются инструменты и методы интеллектуального анализа данных к информации, связанной с образованием [1]. Это направление стало популярным в связи с упрощением получения образовательных данных и увеличением объемов производимой информации различными образовательными учреждениями, онлайн-курсами и т. д.

Интеллектуальный анализ данных (Data Mining) — совокупность методов, которые используются для извлечения новых закономерностей и взаимосвязей из большого набора данных [6]. К методам DM, например, относятся:

- Методы классификации. Данные методы решают задачу разбиения набора данных на определенные группы, называемые классами, на основе одинаковых свойств и признаков. К таким методам можно отнести: метод k -ближайших соседей, деревья решений, метод опорных векторов и т. д.
- Методы кластеризации. Данные методы решают задачу разбиения набора данных на группы, называемые кластерами, на основе похожих признаков. Количество кластеров изначально неизвестно и устанавливается во время работы алгоритма. К таким методам можно отнести: иерархические методы, метод k -средних и т. д.
- Методы регрессии и прогнозирования. Данные методы решают задачу определения зависимости одного параметра от других и предсказания значений параметров на основе выявленных зависимостей.

К стандартным методам интеллектуального анализа данных в EDM добавляются характерные для него методы [1]:

- Обнаружение с помощью моделей (Discovery with Models).
- Обработка данных для принятия решений человеком (Distillation of data for human judgment).

Различия между EDM и DM в первую очередь заключаются в целях анализа. Главными задачами EDM являются улучшение качества образования, повышение эффективности обучения и совершенствование образовательного процесса. Этого можно добиться, например, с помощью корректировки траектории обучения для конкретного студента, прогнозирования успеваемости или формирования персональных рекомендаций [11, 10].

Разница заключается и в используемой информации в области интеллектуального анализа образовательных данных. Собираемые данные зависят от цели анализа и выбранных методов. К тому же эти данные часто имеют сложную структуру и нуждаются в предобработке для применения методов EDM. С данными, используемыми в EDM, тесно связано понятие цифрового следа обучающегося.

Цифровой след — это информация, оставленная в результате активности пользователя при просмотре веб-страниц, хранящаяся в виде файлов cookie [12]. Пользователи могут оставлять цифровой след намеренно или непреднамеренно. Следовательно, понятие цифрового следа делится на два типа:

- Активный цифровой след. Он формируется из информации, оставленной пользователем осознанно. Публикации постов, комментариев и персональных данных, заполнение онлайн-форм — все это составляет активный цифровой след.
- Пассивный цифровой след. Он содержит информацию, собранную заинтересованной стороной без ведома пользователя. Это могут быть данные об активности на сайте, такие как продолжительность нахождения на сайте, количество посещений страниц сайта и переходы между ними.

3.2. Методы EDM

Выше были представлены некоторые методы, используемые в EDM. Они были разделены на несколько категорий [1]. Разберем данные категории более подробно.

3.2.1. Методы прогнозирования

Прогнозирование решает задачу предсказания некоторой переменной на основе выявленных зависимостей. Для выявления зависимостей часть данных должна быть размечена. На основе размеченных данных строится модель, которая в дальнейшем применяется для предсказания новых значений. Основные алгоритмы прогнозирования:

- **Классификация.** Для нового объекта происходит определение значения неизвестного параметра, т. е. класса, на основе выявленных разделяющих признаков. Значение выходного параметра является дискретным.
- **Регрессия.** Отличие от классификации заключается в том, что в данном алгоритме значения выходной переменной представляют собой непрерывные переменные.

Примером использования прогнозирования в EDM может служить задача предсказания успеваемости учащегося [15, 8, 2].

3.2.2. Методы кластеризации

Кластеризация — разбиение анализируемых данных на независимые группы со схожими признаками, т. е. кластеры. Кластеризацию можно назвать неконтролируемой классификацией, т. к. главное отличие этих методов заключается в отсутствии заранее заданного перечня групп.

Кластеризация упрощает задачу анализа данных за счет структурирования информации [9]. Также, применяя методы кластеризации, можно добиться оптимизации анализируемых данных путем хранения только наиболее типичного представителя каждого кластера.

3.2.3. Методы анализирования взаимосвязей

Данные методы направлены на нахождение зависимостей и связей между переменными и их силу. Таким образом, в большом объеме информации, содержащий множество связей, можно определить наиболее важные правила, которые в дальнейшем могут быть использованы для лучшего понимания структуры анализируемых данных или для прогнозирования появления новых событий. В EDM можно встретить применение метода поиска ассоциативных правил [13].

3.2.4. Обнаружение с помощью моделей

Обнаружение с помощью моделей (Discovery with Models) состоит из двух этапов. На первом этапе с помощью методов машинного обучения или человеческого мышления разрабатывается модель. Затем эта модель применяется к данным и используется в качестве компонента в других задачах EDM [4].

3.2.5. Обработка данных для принятия решений человеком

Обработка данных для принятия решений человеком (Distillation of data for human judgment) заключается в представлении данных понятным для человека способом. Можно выделить две цели применения этого метода: классификация и идентификация.

Цель идентификации состоит в отображении данных способом, который позволяет легко определить известные закономерности, а процесс обработки данных для классификации может быть использован при подготовке к процессу построения модели, который значительно ускорит разработку модели прогнозирования.

4. Данные

Источником данных при анализе образовательных данных могут служить лог-файлы учащихся различных образовательных онлайн - платформ. Лог-файл — это файл, содержащий информацию о действиях пользователя. Источником данных послужила платформа «Открытое образование». Набор данных состоит из 828 лог-файлов. Пример записи из лог-файла из набора данных представлен на Рис. 1.

```
{
  "username":"farzonaraim",
  "event_source":"browser",
  "name":"play_video",
  "accept_language":"ru,en;q=0.9",
  "time":"2021-07-12T08:28:47.583464+00:00",
  "agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36 Edg/91.0.864.67",
  "page":"https://courses.openedu.ru/courses/course-v1:ITMOUniversity+DATAN+fall_2020_ITMO/courseware/3c477aa043864248b89820fba9cbc575/c61564b25e7a446eb9dc4c71e7ab925b/?child=first",
  "host":"courses.openedu.ru",
  "session":"ea6b2959475fcff53f3321de8c7c6f9",
  "referer":"https://courses.openedu.ru/courses/course-v1:ITMOUniversity+DATAN+fall_2020_ITMO/courseware/3c477aa043864248b89820fba9cbc575/c61564b25e7a446eb9dc4c71e7ab925b/?child=first",
  "context":{
    "user_id":534023,
    "org_id":"ITMOUniversity",
    "course_id":"course-v1:ITMOUniversity+DATAN+fall_2020_ITMO",
    "path":"/event"
  },
  "ip":"176.59.22.126",
  "event":{
    "duration":"624.520000",
    "code":"html5",
    "id":"1f74119c2c55441e8b4a29df10277f66",
    "currentTime":6.66351
  },
  "event_type":"play_video"
}
```

Рис. 1: Пример лог-записи

Каждая запись включает в себя общие поля, содержащие ключевую информацию о событии и о пользователе, вызвавшем его. Можно выделить поля, которые хранят следующую информацию:

- идентификатор пользователя (*user_id* в лог-записи);
- имя пользователя (*username* в лог-записи);
- идентификатор курса (*course_id* в лог-записи);
- тип события (*event_type* в лог-записи);

- время события (*time* в лог-записи).

Эти данные содержатся в каждой лог-записи, независимо от типа вызванного события.

4.1. Группы событий

В зависимости от действий пользователя генерируются разные типы событий. Например, взаимодействие с заданиями инициирует тип события *problem_check*, который сообщает об успешной проверке задания.

Для анализа онлайн-курсов не все типы событий представляют интерес. Значимой информацией в данном случае являются данные, которые могут помочь в исследовании эффективности курса и оценке его сложности. Нужные события можно разделить на несколько групп:

- события взаимодействия с видео;
- события взаимодействия с заданиями;
- события взаимодействия с учебниками;
- навигационные события;
- события регистрации на курсе;
- события окончания курса.

Для каждого типа события определена специфическая конкретно для этого события информация.

4.1.1. События регистрации на курсе

События, генерирующиеся при зачислении студента на курс.

Тип события: *edx.course.enrollment.activated*

Поля:

Поле	Описание
<i>course_id</i>	идентификатор курса, на который был зачислен студент
<i>user_id</i>	идентификатор зачисленного студента

Таблица 1: Описание полей для события типа *edx.course.enrollment.activated*

4.1.2. Навигационные события

События, генерирующиеся при передвижении студента по курсу.

Типы событий:

- *seq_goto* — генерируется при переходе студента между блоками последовательности.
- *seq_next* — генерируется при переходе студента к следующему блоку последовательности.
- *seq_prev* — генерируется при переходе студента к предыдущему блоку последовательности.

Поля:

Поле	Описание
<i>id</i>	идентификатор последовательности
<i>new</i>	индекс блока, к которому осуществляется переход
<i>old</i>	индекс блока, с которого осуществляется переход

Таблица 2: Описание полей для событий типа *seq_goto*, *seq_next* и *seq_prev*

4.1.3. События взаимодействия с видео

События, генерирующиеся при взаимодействии студента с видео.

Типы событий:

- *pause_video* — генерируется при остановке видео.

- *play_video* — генерируется при запуске видео.
- *seek_video* — генерируется при переходе к другому месту в видео.
- *speed_change_video* — генерируется при изменении скорости видео.

Поля:

Поле	Описание
<i>id</i>	идентификатор видео
<i>current_time</i>	для <i>pause_video</i> , <i>play_video</i> и <i>speed_change_video</i> время в видео, когда произошло событие
<i>new_time</i>	для <i>seek_video</i> время видео, к которому переходит пользователь
<i>old_time</i>	для <i>seek_video</i> время видео, из которого переходит пользователь
<i>new_speed</i>	для <i>speed_change_video</i> новая выбранная скорость видео
<i>old_speed</i>	для <i>speed_change_video</i> скорость, с которой воспроизводилось видео

Таблица 3: Описание полей для событий типа *pause_video*, *play_video*, *seek_video* и *speed_change_video*

4.1.4. События взаимодействия с заданиями

События, генерирующиеся при проверки ответа на задания.

Тип события: *problem_check*

Поля:

Поле	Описание
<i>problem_id</i>	идентификатор задания
<i>attempts</i>	количество попыток
<i>grade</i>	полученная оценка
<i>max_grade</i>	максимально возможная оценка
<i>success</i>	правильность выполнения задания

Таблица 4: Описание полей для события типа *problem_check*

4.1.5. События взаимодействия с учебниками

События, генерирующиеся при передвижении студента в средстве просмотра учебника.

Типы событий:

- *gotopage* — генерируется при переходе к определенной странице в документе.
- *prevpage* — генерируется при переходе к предыдущей странице в документе.
- *nextpage* — генерируется при переходе к следующей странице в документе.

Поля:

Поле	Описание
<i>chapter</i>	имя файла
<i>new</i>	страница, на которую был совершен переход
<i>old</i>	для <i>gotopage</i> страница, с которой был совершен переход

Таблица 5: Описание полей для событий типа *gotopage*, *prevpage* и *nextpage*

4.1.6. События окончания курса

События, генерирующиеся при получении сертификата студентом по окончании курса.

Тип события: *edx.certificate.created*

Поля:

Поле	Описание
<i>course_id</i>	идентификатор курса, для которого выдан сертификат
<i>user_id</i>	идентификатор студента, получившего сертификат

Таблица 6: Описание полей для события типа *edx.certificate.created*

4.2. База данных

В качестве СУБД была выбрана PostgreSQL. Схема базы данных представлена на Рис. 2. Взаимодействие с базой данных осуществлялось с помощью ORM библиотеки рееее языка программирования Python. После обработки логов было получено 6412 уникальных пользователей и 18 онлайн-курсов.

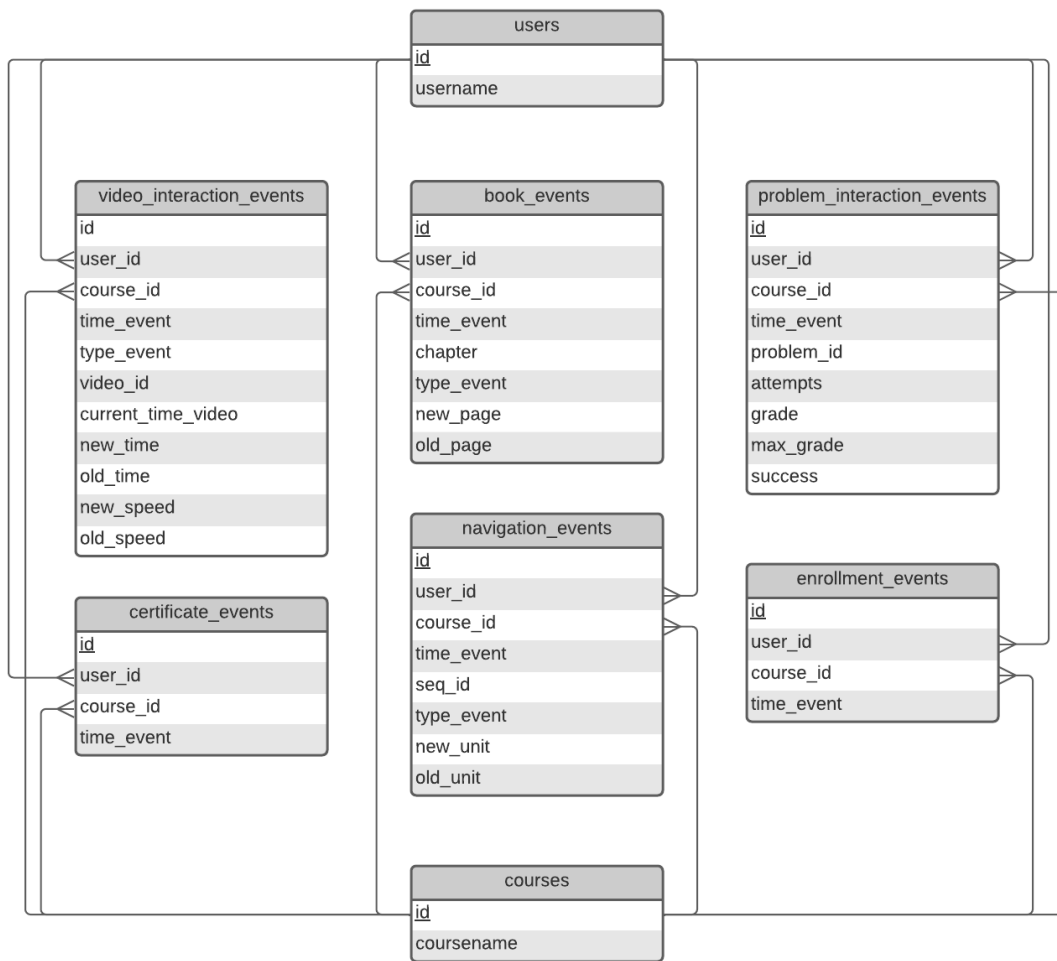


Рис. 2: Схема базы данных

5. Решение

5.1. Архитектура решения

На Рис. 3 представлена схема взаимодействия модулей. Необходимая информация из лог-файлов загружается в базу данных с помощью модуля обработки лог-файлов. Взаимодействие с базой данных происходит посредством ORM реееее языка программирования Python. Далее отображается структурированная информация о курсе и о найденных кластерах студентов для анализа онлайн-курса и определения шаблонов поведения.

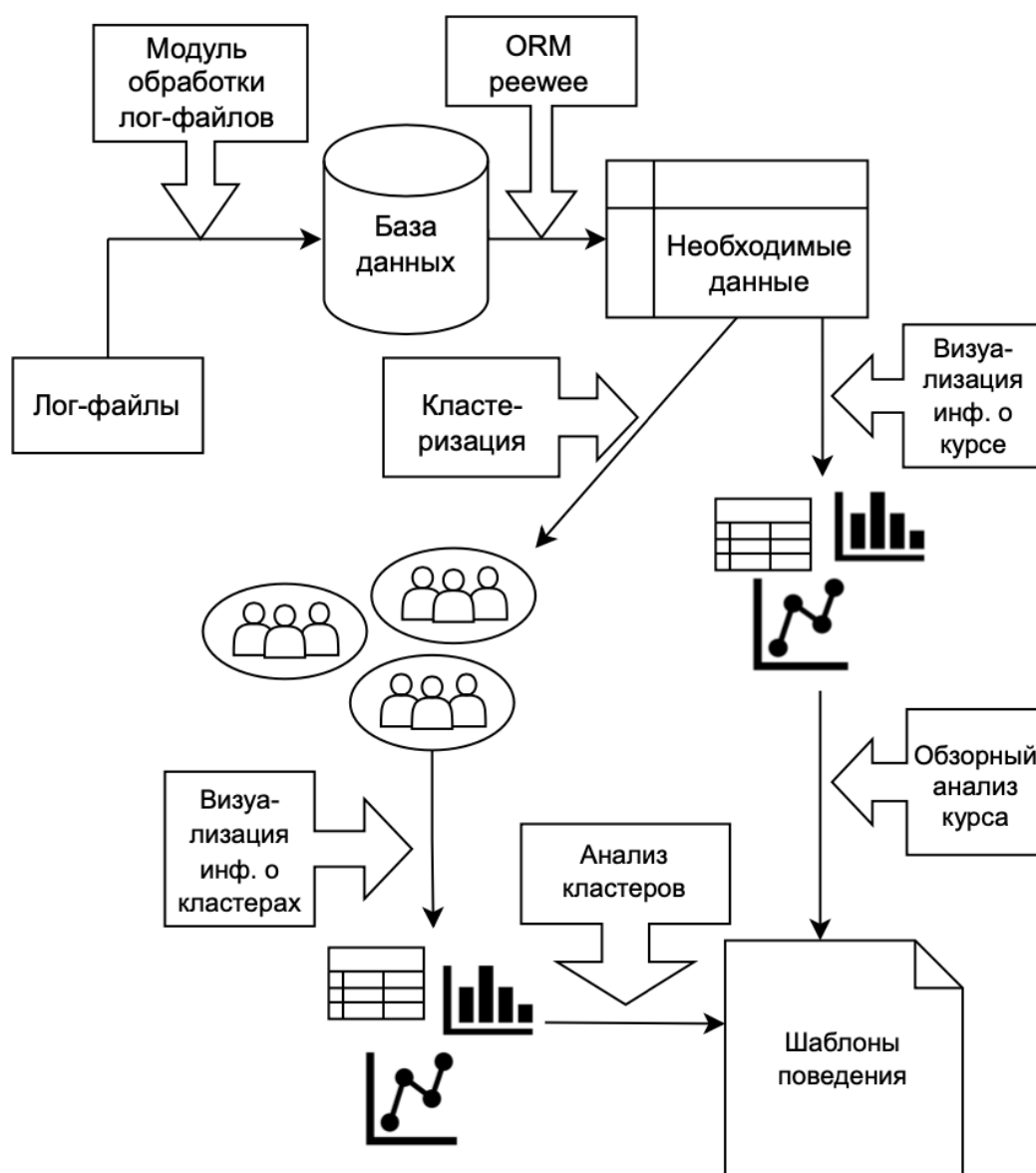


Рис. 3: Схема взаимодействия модулей

5.2. Обработка логов

Для обработки логов была создана процедура на языке Python. Процедура на вход получает лог-файл, затем обрабатывает каждую запись в лог-файле, определяя тип события в лог-записи и извлекая оттуда соответствующую данному типу события информацию. Обработанные данные загружаются в базу данных посредством использования библиотеки `reewee`.

5.3. Анализ онлайн-курса

Для вывода информации о курсе была создана процедура, которая на вход получает данные анализируемого онлайн-курса. В ходе работы процедуры устанавливаются студенты, замеченные в возможно нечестном прохождении курса. Их действия не учитываются при исследовании онлайн-курса. Примеры вывода процедуры показаны на Рис. 4 и Рис. 5. Результатом работы данной процедуры являются графики, показывающие следующую информацию:

- процент выполненных заданий студентом;
- количество удачных/неудачных попыток выполнения каждого задания;
- активность на курсе в течение всего времени;
- среднюю активность учащихся в течение дня;
- среднее время участия студента на курсе;
- количество обращений к материалам курса.

Также выводится информация о студентах, подозреваемых в списывании при прохождении онлайн-курса (данные студенты определяются посредством анализа количества затраченных попыток на прохождение заданий) и о заданиях, возможно составленных с ошибкой (данные задания определяются с помощью анализа общего количества

удачных и неудачных попыток выполнения задания всеми студентами курса).

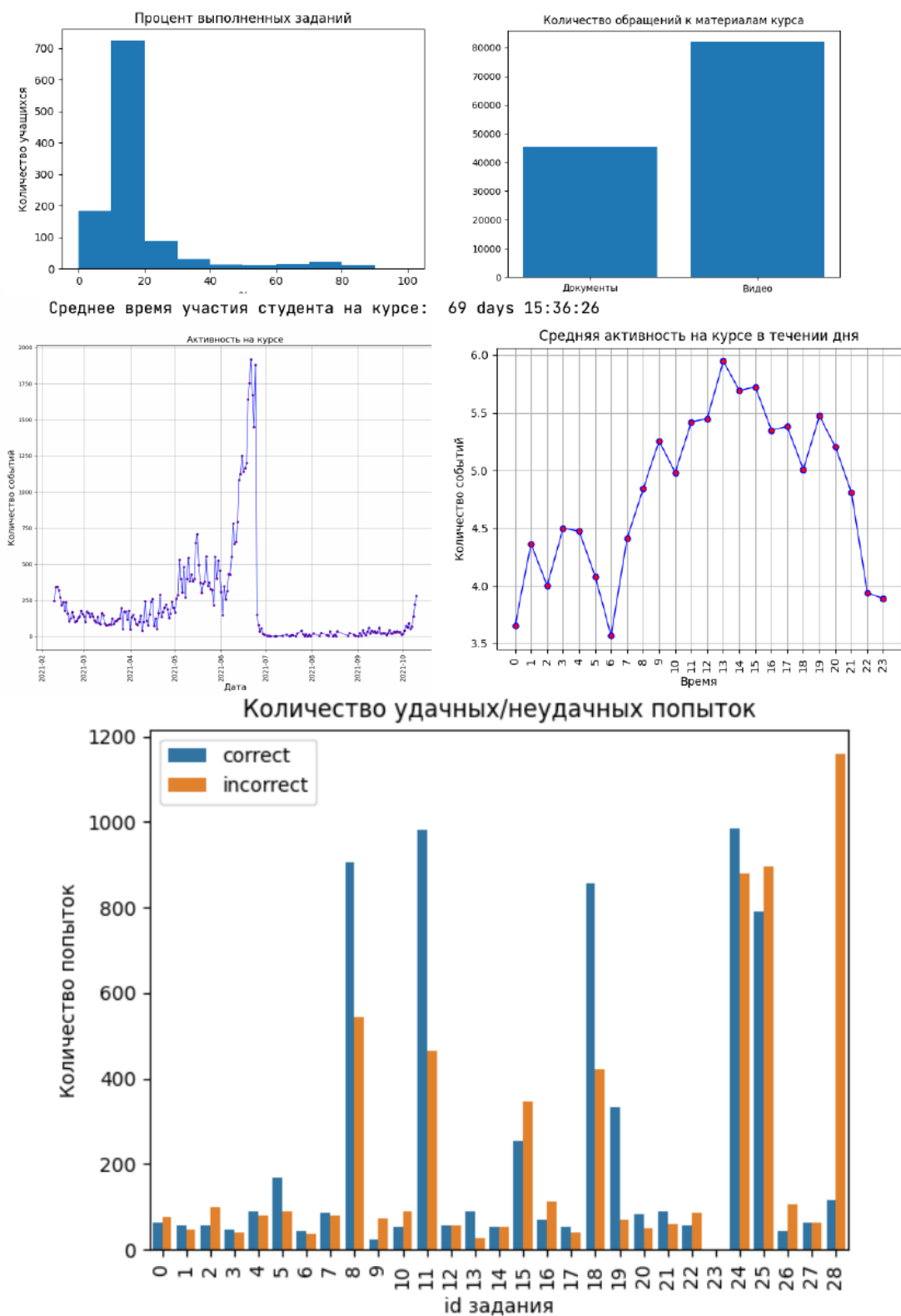


Рис. 4: Пример результата работы процедуры по анализу онлайн-курса



Рис. 5: Пример результата работы процедуры по анализу онлайн-курса

5.4. Кластеризация

Для определения кластеров студентов была создана процедура, которая на вход получает данные анализируемого онлайн-курса. Множество признаков содержит:

- количество взаимодействий с курсом в течение каждого периода;
- количество попыток, потраченных на выполнение каждого задания;
- количество взаимодействий с видео и с текстовыми материалами;
- процент успешности выполнения курса.

Так как данные признаки измеряются в разных единицах используем стандартизацию (StandardScaler из библиотеки sklearn). В качестве кластеризации был выбран иерархический подход из библиотеки scipy. Евклидово расстояние и метод Варда были выбраны в качестве меры близости между объектами и метода кластеризации соответственно. Проиллюстрируем работу процедуры на примере.

Для выбора оптимального количества кластеров строятся дендрограмма и график, иллюстрирующий метод «локтя» (представлены на Рис. 6). На их основе можно выбрать оптимальное количество кластеров или максимально возможное расстояние между кластерами.

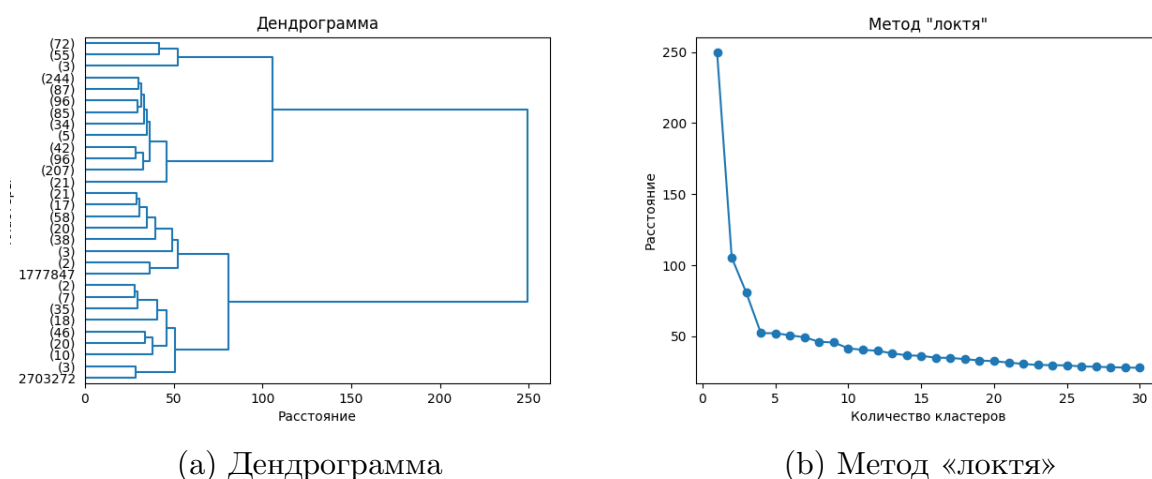


Рис. 6: Выбор оптимального количества кластеров

Выберем оптимальное количество кластеров равное 2 и проинтерпретируем полученные кластеры, построив некоторые графики.

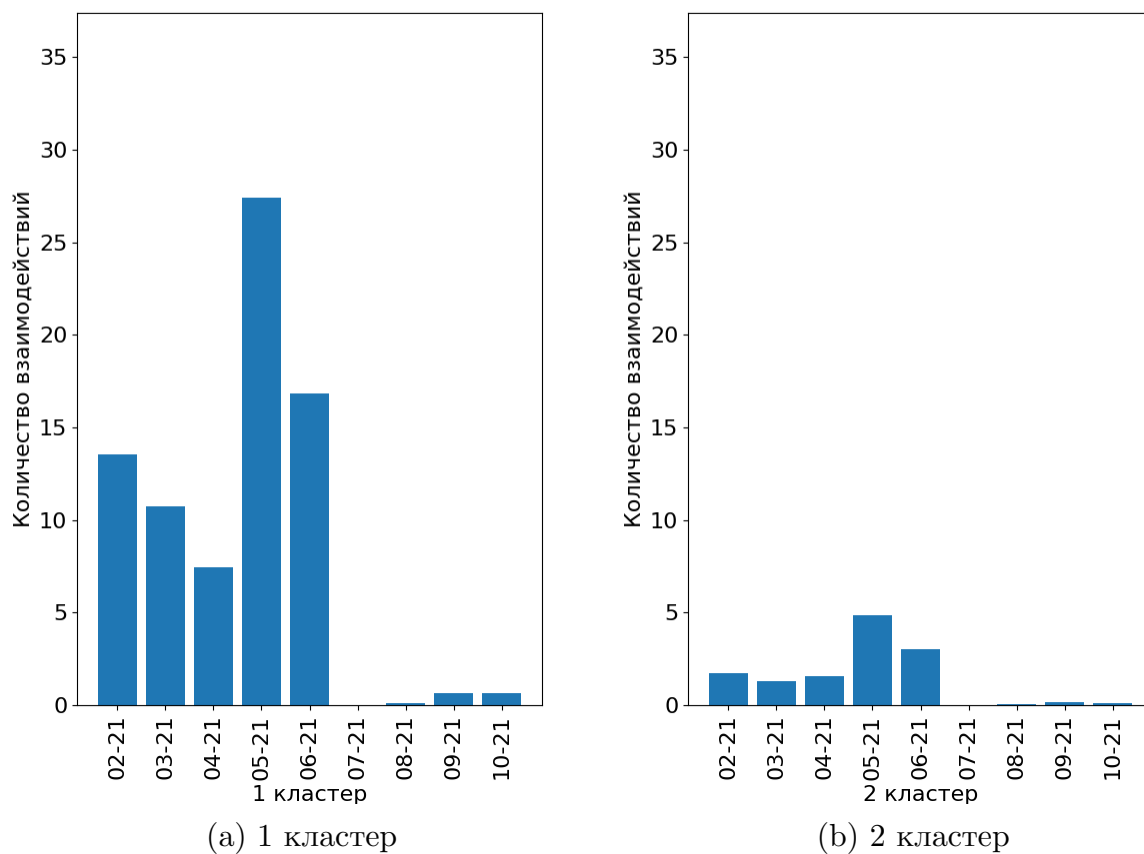


Рис. 7: Среднее число взаимодействий с курсом по кластерам

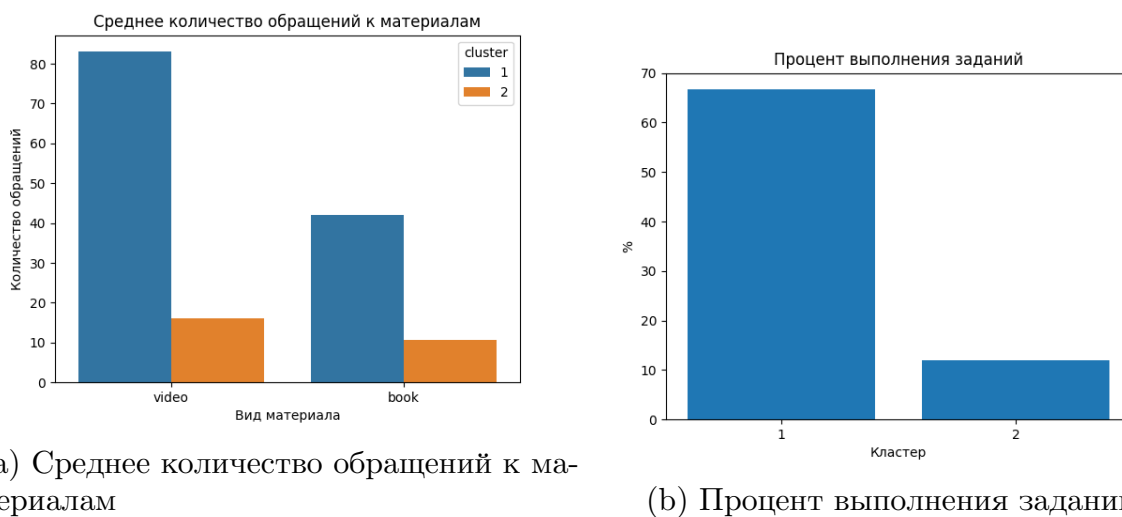


Рис. 8: Графики среднего количества обращений к материалам и процент выполнения заданий



(a) 1 кластер



(b) 2 кластер

Рис. 9: Среднее число попыток выполнения заданий по кластерам

По графикам на Рис. 7 видно, что студенты первого кластера взаимодействовали с курсом гораздо больше студентов второго кластера. Пик взаимодействия пришелся на май 2021 года и в первом, и во втором кластерах. На графике на Рис. 8а можно также увидеть, что количество обращений к материалам студентами первого кластера больше, чем студентами второго кластера. Можно заметить, что большее предпочтение студенты обоих кластеров отдают видеоматериалам. Рис. 8б демонстрирует, что успешность прохождения курса студентами первого кластера значительно выше, чем второго. Исходя из графиков, представленных на Рис. 9, прослеживается, что студенты первого кластера тратили на задания большее количество попыток, чем студенты второго кластера. Также по этим графикам можно отследить самые сложные

и самые легкие задания курса.

Проинтерпретируем полученные кластеры. Можно предположить, что во второй кластер попали студенты, которые мало взаимодействовали с материалами курса и со всем курсом в целом, что привело к низкой успеваемости. В первый кластер попали студенты, относящиеся к прохождению онлайн-курса серьезнее. Это можно увидеть по большему количеству взаимодействий с курсом и учебными материалами. Такое поведение привело к улучшению успеваемости по курсу.

Приведем результаты кластеризации на примере другого онлайн-курса. Выберем оптимальное количество кластеров равное 4 (дендрограмма и график, иллюстрирующий метод «локтя» (представлены на Рис. 10)

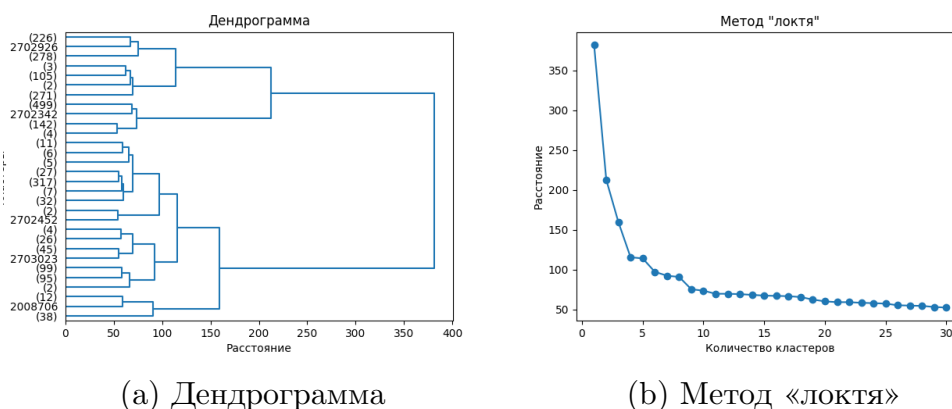


Рис. 10: Выбор оптимального количества кластеров

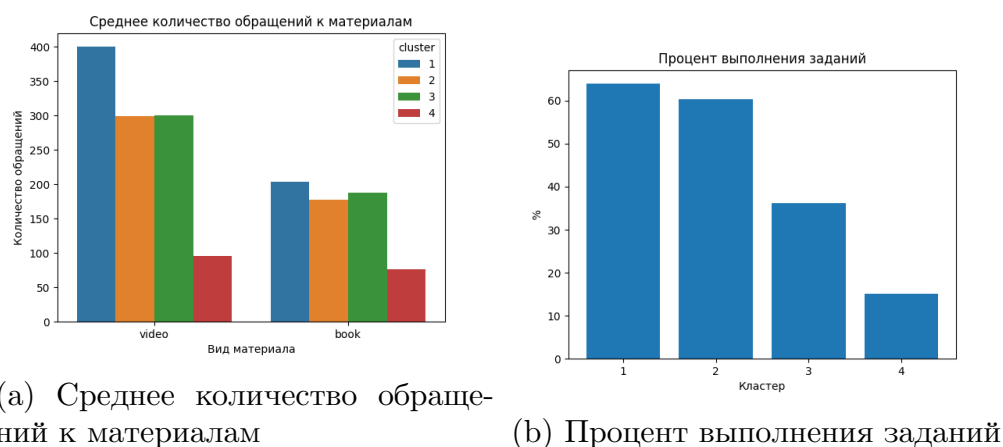
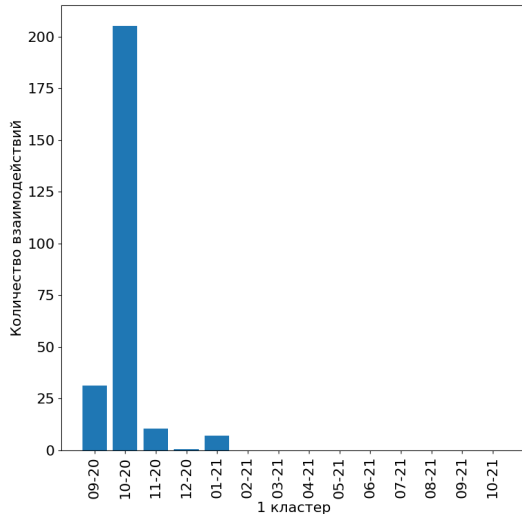
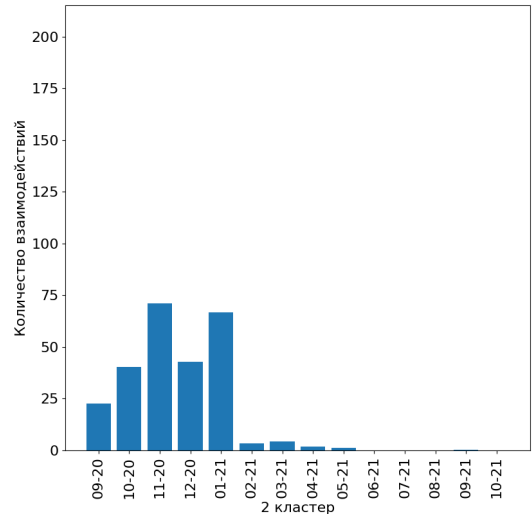


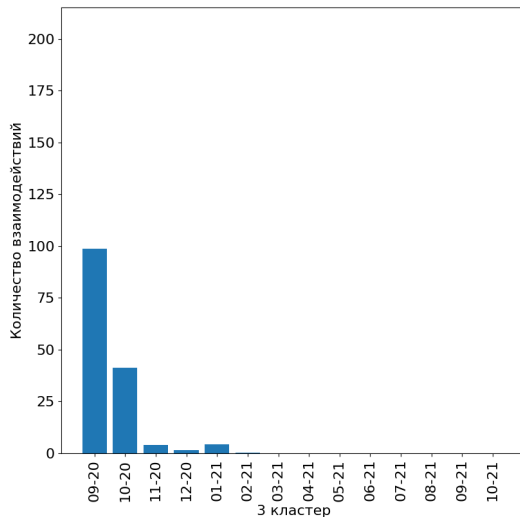
Рис. 11: Графики среднего количества обращений к материалам и процент выполнения заданий



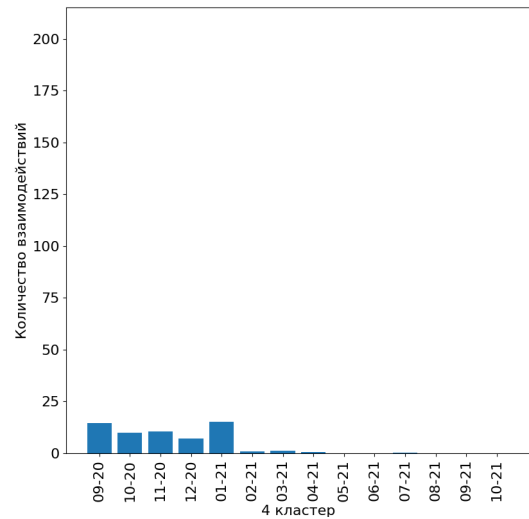
(a) 1 кластер



(b) 2 кластер



(c) 3 кластер



(d) 4 кластер

Рис. 12: Среднее число взаимодействий с курсом по кластерам



(a) 1 кластер



(b) 2 кластер



(c) 3 кластер



(d) 4 кластер

Рис. 13: Среднее число попыток выполнения заданий по кластерам

По графикам на Рис. 12 видно, что количество взаимодействий с курсом студентов первого кластера наравне со студентами второго кластера, однако взаимодействия студентов первого кластера в большей степени пришлись на октябрь 2020 года, тогда как студенты второго кластера распределили нагрузку более равномерно. Студенты третьего кластера, как и студенты первого кластера, проходили онлайн-курс преимущественно в начале учебного года, однако их количество взаимодействий меньше. Студенты четвертого кластера мало взаимодействовали с курсом в течение всех месяцев. На графике на Рис. 11а можно увидеть, что большее количество обращений к материалам у студентов первого кластера. Студенты второго и третьего кластеров обращались к материалам в примерно равных количествах, а студенты четвертого кластера использовали материалы меньше всех. Также можно проследить, что студенты всех кластеров отдавали предпочтение видеоматериалам. Рис. 11b демонстрирует, что успешность прохождения курса студентами первого и второго кластеров значительно выше, чем третьего и четвертого. Исходя из графиков, представленных на Рис. 13, прослеживается, что студенты первого и второго кластеров тратили на задания большее количество попыток, чем студенты остальных кластеров. Также по этим графикам можно отследить самые сложные и самые легкие задания курса.

Проинтерпретируем полученные кластеры. Можно предположить, что в первый и второй кластеры попали самые прилежные учащиеся. У этих студентов практически одинаковая успеваемость, различие лишь в равномерности прохождения онлайн-курса. Студенты третьего и четвертого кластеров имеют успеваемость ниже половины и малое количество взаимодействий с курсом. Однако по количеству взаимодействий с материалами можно сделать вывод, что студенты 3 кластера пытались разобраться с предложенной онлайн-курсом теорией, тогда как студенты 4 класса не были заинтересованы в надлежащем прохождении курса.

6. Заключение

В рамках данной работы был выполнен анализ цифрового следа для оптимизации и оценки результатов онлайн-обучения.

В результате были решены следующие задачи:

- Изучена предметная область EDM.
- Рассмотрены существующие подходы в EDM.
- Определена значимая информация в логах.
- Спроектирована схема базы данных.
- Реализованы инструменты для обработки логов и их загрузки в базу данных.
- Разработаны инструменты для анализа онлайн-курса и определения шаблонов поведения учащихся.

Код доступен по ссылке [github](#).

Список литературы

- [1] Baker Ryan. Data mining for education // International Encyclopedia of Education. — 2010. — 01. — Vol. 7. — P. 112–118.
- [2] Brahim Ghassen Ben. Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features // [Arabian Journal for Science and Engineering](#). — 2022. — URL: <https://doi.org/10.1007/s13369-021-06548-w>.
- [3] Data Mining and Curriculum Development in Higher Education / Mirela Mabić, Fuad Dedić, Nina Bijedic, Dražena Gašpar. — 2017. — 06.
- [4] Discovery With Models / Arnon HersHKovitz, Ryan Baker, Janice Gobert et al. // [American Behavioral Scientist](#). — 2013. — 10. — Vol. 57. — P. 1480–1499.
- [5] Educational Data Mining and Personalized Support in Online Introductory Physics Courses / Farook Mohammad Al-Shamali, Hongxin Yan, Sabine Graf, Fuhua Oscar Lin // EDM. — 2020.
- [6] Hung Jui-Long, Rice Kerry, Saba Anthony. An Educational Data Mining Model for Online Teaching and Learning // [Journal of Educational Technology Development and Exchange](#). — 2012. — 06. — Vol. 5. — P. 77–94.
- [7] Kamalov Firuz, Sulieman Hana, Santandreu Calonge David. Machine learning based approach to exam cheating detection // [PLOS ONE](#). — 2021. — 08. — Vol. 16, no. 8. — P. 1–15. — URL: <https://doi.org/10.1371/journal.pone.0254340>.
- [8] Kaur Parneet, Singh Manpreet, Josan Gurpreet. Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector // [Procedia Computer Science](#). — 2015. — 12. — Vol. 57. — P. 500–508.

- [9] Križanić Snježana. Educational data mining using cluster analysis and decision tree technique: A case study // [International Journal of Engineering Business Management](#). — 2020. — Vol. 12. — P. 1847979020908675. — <https://doi.org/10.1177/1847979020908675>.
- [10] Magdin Martin, Turčáni M. Personalization of student in course management systems on the basis using method of data mining // [Turkish Online Journal of Educational Technology](#). — 2015. — 01. — Vol. 14. — P. 58–67.
- [11] Malvandi Shahrbanoo, Farahi Ahmad. Provide a Method for Increasing the Efficiency of Learning Management Systems using Educational Data Mining // [Indian Journal of Science and Technology](#). — 2015. — 10. — Vol. 8.
- [12] McDermot Maureen. Digital footprints: Creation, implication, and higher education. — 2018.
- [13] Mining Rare Association Rules from e-Learning Data / Cristóbal Romero, José Raúl Romero, José María Luna, Sebastian Ventura. — 2010. — 10. — P. 171–180.
- [14] Okike Ezekiel Uzor, Mogorosi Merapelo. Educational Data Mining for Monitoring and Improving Academic Performance at University Levels // [International Journal of Advanced Computer Science and Applications](#). — 2020. — Vol. 11.
- [15] Pal Saurabh. Mining Educational Data Using Classification to Decrease Dropout Rate of Students // [International Journal of Multidisciplinary Sciences and Engineering](#). — 2012. — 05. — Vol. 3. — P. 35–39.
- [16] Using Educational Data Mining to Identify and Analyze Student Learning Strategies in an Online Flipped Classroom / Randall Davies, Gove Allen, Conan Albrecht et al. // [Education Sciences](#). — 2021. — Vol. 11, no. 11. — URL: <https://www.mdpi.com/2227-7102/11/11/668>.