

ИСТОРИЧЕСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М. В. ЛОМОНОСОВА  
АССОЦИАЦИЯ «ИСТОРИЯ И КОМПЬЮТЕР»

---

# **ИСТОРИЧЕСКИЕ ИССЛЕДОВАНИЯ В КОНТЕКСТЕ НАУКИ О ДАННЫХ:**

**Информационные ресурсы,  
аналитические методы  
и цифровые технологии**

Материалы международной конференции

Москва, 4–6 декабря 2020 г.



---

Москва – 2020

УДК 930.2  
ББК 63ф1я431  
И90

*Конференция поддержана Российским фондом фундаментальных исследований,  
грант 20-09-22024\20*

Ответственный редактор: *И. М. Гарскова*

Редколлегия:

чл.-корр РАН, д.и.н. *Л. И. Бородин*, д.и.н. *В. Н. Владимиров*,  
к.и.н. *И. М. Гарскова*

Рецензенты:

чл.-корр. РАН *Н. М. Арсентьев*, д.т.н. *Г. З. Залаев*

Редакционный совет:

д.и.н. *С.А. Баканов* (Челябинск), д.и.н. *Ш. Д. Батырбаева* (Бишкек),  
д.и.н. *С. И. Корниенко* (Пермь), д.и.н. *Л. Н. Мазур* (Екатеринбург),  
д.и.н. *Ю. Ю. Юмашева* (Москва)

**Исторические исследования в контексте науки о данных:  
И90 информационные ресурсы, аналитические методы и цифро-  
вые технологии.** Материалы международной конференции. Мо-  
сква, 4–6 декабря 2020 г. – Москва : МАКС Пресс, 2020. – 484 с.  
ISBN 978-5-317-06529-4

<https://doi.org/10.29003/m1781.978-5-317-06529-4>

Сборник содержит тексты докладов Международной конференции.  
В рамках указанной международной конференции была проведена XVII  
конференция Межрегиональной Ассоциации «История и компьютер».

*Ключевые слова:* история, историческая информатика, наука о дан-  
ных, методы исследования, информационные ресурсы, цифровые тех-  
нологии.

УДК 930.2  
ББК 63ф1я431

ISBN 978-5-317-06529-4

© Ассоциация «История и компьютер», 2020  
© Исторический факультет МГУ имени  
М. В. Ломоносова, 2020  
© Оформление. ООО «МАКС Пресс», 2020

<https://doi.org/10.29003/m1847.978-5-317-06529-4/447-453>

**Оськин Д.А.**

*Белорусский государственный экономический университет,  
Минск*

**Оськин А.Ф.**

*Полоцкий государственный университет*

### **Образовательная аналитика.**

#### **Новый подход к повышению качества учебного процесса**

**Аннотация.** В статье описывается тенденции в области онлайн образования, вызванные пандемией COVID-19. Обосновывается внедрение образовательной аналитики в образовательный процесс. Рассматриваются основные методы и инструменты образовательной аналитики. На конкретном примере разбирается построение и оценка модели классификации студентов с использованием высокоуровневого языка программирования Python.

**Ключевые слова:** Дистанционное обучение, образовательная аналитика, машинное обучение, Python.

**Oskin D.A.**

*Belarusian State Economic University, Minsk*

**Oskin A.F.**

*Polotsk State University*

### **Educational analytics.**

#### **A new approach to improving the quality of the educational process**

**Abstract.** This article describes the trends in online education caused by the COVID-19 pandemic. The introduction of learning analytics into the educational process is substantiated. The main methods and tools of educational analytics are considered. Using a specific example, we will understand the construction and assessment of a student classification model using the high-level programming language Python.

**Keywords:** Distance learning, learning analytics, machine learning, Python.

### **Введение**

С началом пандемии вызванной COVID-19 учебные заведения всех уровней обучения, по всему миру частично или полностью перевели процесс обучения в онлайн. По данным ЮНЕСКО 53 государства уже развернули национальные образовательные платформы для дистанционного обучения[1]. При этом активно используются такие методики преподавания как смешанное и дистанционное обучение. Кроме этого наблюдается взрывной рост интереса к онлайн-курсам размещенных на MOOCплатформах. Согласно отчету, опубликованному на ресурсе ClassCentralMOOCReport [2], количество

новых пользователей зарегистрированных в апреле 2020 года на трех наиболее популярных платформах (Coursera, edX, FutureLearn) превысило суммарное количество новых пользователей за весь 2019 год. По мнению экспертов Всемирного банка, резкий переход на дистанционное образование вызовет ряд сложностей, а именно [3, стр. 4–6]:

- На данный момент очень мало систем образования (даже среди самых высококлассных), которые обладают хорошим техническим обеспечением, чтобы осуществить быстрый переход на дистанционное обучение.

- Переход на дистанционное обучение требует огромных затрат. Безусловно, важно обеспечить инфраструктурный потенциал

- Переход на дистанционное обучение вызывает обеспокоенность в связи с социальным расслоением.

- Большинство учащихся будет испытывать трудности при переходе к дистанционному образованию.

- При переходе на дистанционное обучение нужно иметь в виду, что поначалу учащиеся будут показывать более слабые результаты.

- Перед учебными заведениями встанет вопрос выбора, каким предметам обучать онлайн, а какие оставить для самостоятельного освоения.

- Только некоторые учителя будут способны осуществить быстрый и эффективный переход к онлайн-модели обучения, поскольку процессы преподавания удаленно и в классах значительно отличаются друг от друга.

Помимо этого, мы можем говорить о том, что переход на дистанционные формы обучения повлечет за собой увеличение количества отчисленных студентов. Так, по данным исследования [4], показатели завершения курсов для некоторых групп студентов, обучающихся дистанционно, могут быть на 22% ниже, чем для студентов, обучающихся на очных курсах. В тоже время средний процент завершения МООС курсов составляет не больше 10% от общего количества записавшихся на прохождение курса [5].

Одним из возможных путей улучшения ситуации с дистанционным обучением является объективный анализ процесса обучения путем применения образовательной аналитики. Это позволит выявить трудности, связанные с адаптацией студентов и пониманием их поведения и увеличит количество успешных студентов.

### **Образовательная аналитика: определения, цели, задачи**

Образовательная аналитика относительно новая сфера научной и практической деятельности. Так, профессиональное сообщество ученых и практиков Society of Learning Analytics Research (SoLAR) образовано в 2010 году, а первая конференция под эгидой сообщества прошла в 2011 году.

Согласно определению, предложенному Джорджем Сименсом на первой конференции под эгидой SoLAR, образовательная аналитика – это измерение, сбор, анализ и предоставление данных об учащих и других сопутствующих факторах, с целью понимания и оптимизации обучения и условий, в которых оно происходит [6].

Цель образовательной аналитики – дать оценку поведению пользователей в контексте преподавания и обучения с ее последующими анализом и интерпретацией. Это позволяет получить новую информацию об учебном процессе и его участниках вместе с новыми моделями улучшения обучения, преподавания, эффективной организации учебного процесса, принятия решений [6]. Ключевым является факт передачи полученных знаний педагогам и студентам для оптимизации их педагогического и образовательного поведения, с целью развития соответствующих навыков в данной области, лучшего понимания образовательного процесса и сопутствующих ему областей, например, таких как университетский маркетинг и бизнес.

Основная задача образовательной аналитики – сбор и анализ данных, а также проведение своевременных вмешательств, с целью совершенствования процесса обучения у обучающихся [7].

Образовательная аналитика фокусируется на образовательном процессе, с результирующими преимуществами для всех основных участников процесса [8]. Romero and Ventura (2013) разделили основных участников в соответствии с их основными целями, на четыре группы (см. табл. 1).

Таким образом, образовательная аналитика это:

- подход, в основе которого лежит анализ данных;
- подход, затрагивающий всех участников образовательного процесса и все структурные подразделения учреждения высшего образования, ответственные за образовательный процесс;
- процесс, для организации которого требуется интеграция информационных потоков из различных источников, относящихся к поддержке и проведению образовательного процесса в организации;
- процесс, для успешной реализации которого требуется понимание подходов к анализу данных, наличие ресурсов для его реализации, правильная постановка целей процесса.

Таблица 1

**Цели основных участников образовательного процесса**

| Участник процесса        | Основная цель, преимущества, перспективы                                                                                                                     |
|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Обучающиеся              | Поддержка обучающихся с помощью адаптивных форм обратной связи, рекомендаций, отвечающих индивидуальным потребностям, с целью улучшения показателей обучения |
| Преподаватели (Educator) | Понимание процесса обучения студентов, отражение образовательных методик и показателей, когнитивные и бихевиористские аспекты                                |
| Аналитики                | Подбор и применение техник образовательной аналитики согласно текущей проблеме, оценка эффективности обучения в различных ситуациях.                         |
| Администраторы           | Оценка ресурсов и их образовательной стоимости                                                                                                               |

Справедливость этих выводов подтверждается опубликованным в 2015 году исследованием Improving the Quality and Productivity of the Higher Education Sector [9] от сообщества SoLAR. В работе проанализирован опыт внедрения образовательной аналитики в вузах США, Австралии и Великобритании. В ходе анализа выявлено, что основными целями внедрения образовательной аналитики, как правило, являются: повышение качества образования, эффективности университета и снижение расходов, а также удержание студентов, особенно на начальном этапе обучения. Инструментом реализации целей и задач являются рабочие группы, службы или институты, создаваемые внутри университетов. Помимо этого, практически у всех вузов реализация осуществляется через пилотные проекты по анализу данных.

**Методы и инструменты образовательной аналитики**

Поскольку основой для образовательной аналитики являются данные, то используемые методы и инструменты характерны для анализа данных в целом. К наиболее часто встречающимся методам, по мнению авторов [10], относят следующие.

## 1. Прогнозные модели:

- классификация;
- регрессионный анализ;
- оценка скрытых знаний.

2. Обнаружение структуры:
  - кластеризация;
  - факторный анализ.
3. Поиск закономерностей (взаимосвязей)
  - поиск по ассоциативным правилам;
  - поиск по последовательной схеме;
  - корреляционный анализ;
  - поиск причинных взаимосвязей.
4. Открытие с помощью модели.

Если говорить об инструментах, наиболее часто применяемых для реализации методов образовательной аналитики, то на основании опроса, проведенного на платформе сообщества аналитиков данных Kaggle в 2018 году, можно выделить два наиболее популярных языка программирования для интеллектуального анализа данных – Python и R [11].

Согласно опросу обнаружено, что 83% из почти 24 000 профессионалов в области анализа данных используют Python. Специалисты по обработке данных и программисты любят Python, за его универсальность. Python оказывается быстрее R в циклах с менее 1000 итераций. Специалисты отмечают, что Python лучше R для предобработки данных. Кроме этого в Python встроены пакеты для обработки естественного языка и анализа данных, а также язык является объектно-ориентированным. В тоже время R лучше подходит для специального анализа и изучения наборов данных, чем Python. Это язык с открытым исходным кодом и программное обеспечение для статистических вычислений и графики. Это непростой язык для изучения, и большинство людей считают, что с Python легче освоиться.

В циклах, которые имеют более 1000 итераций, R фактически превосходит Python. Это может заставить некоторых задуматься, лучше ли R для анализа данных с большими наборами данных, однако R был создан статистиками, что и отражено в его операциях. Приложения для обработки данных в Python кажутся более естественными.

Таким образом, можно сделать вывод, что на текущий момент наиболее популярным языком для реализации методов образовательной аналитики является язык Python.

### Пример из практики

Рассмотрим на практическом примере применение образовательной аналитики в образовательном процессе учреждения высшего образования. Начиная с 2017 учебного года в Высшей школе управления и бизнеса при Белорусском государственном экономическом университете ведется прием студентов на дистанционную форму обучения по специальности «Правоведение» (специализация «Хозяйственное право»). В качестве системы управления обучением (СУО) выбрана платформа Moodle.

Цель исследования – снижение количества отчисленных студентов после первого семестра обучения. Задача исследования – разработка и внедрение модели раннего выявления студентов, подверженных высокому риску отчисления на основе их взаимодействия с системой управления обучением. На момент построения модели на специальность суммарно зачислены 238 студентов, 67 из которых в 2019 году.

Данные для построения модели – таблица с выгрузкой активностей всех студентов специальности «Правоведение» с сентября 2017 по февраль 2020 года в формате csv. В качестве инструмента выбран язык программирования Python 3.7 с подключенными программной библиотекой для обработки и анализа данных Pandas и библиотекой машинного обучения Scikit-learn. Средой реализации модели выбрано бесплатное веб-приложение с открытым исходным кодом Jupiter Notebook. Временной интервал для прогнозирования 60 дней с момента первого взаимодействия студента с СУО. Т.к. задача решается методом классификации, то в качестве метода для модели выбран алгоритм машинного обучения «Дерево решений».

На первом этапе построения модели проводится подготовка данных к анализу. В рассматриваемом случае остаются лишь столбцы, непосредственно описывающие взаимодействие студентов с модулями СУО. Далее студенты размечаются по признаку продолжения учебы после первого семестра на тех, кто продолжил учиться – группа 1 и тех, кто отчислен – группа 0. После разметки создается сводная таблица по количеству действий студентов с курсами в рамках первого семестра, количеству уникальных дней пребывания в системе и дней учебы. После этого таблица разбивается на обучающую – студенты, зачисленные до 2019 года и тестовую выборку – студенты, зачисленные в 2019 году. Из тестовой выборки убирается столбец с разметкой по признаку – учится/отчислен.

Таблица 2

Таблица представлений обучающей выборки

| userid | course_tried | study_days | reviewed | started | submitted | viewed | stay_studied |
|--------|--------------|------------|----------|---------|-----------|--------|--------------|
| 88     | 1            | 12         | 15       | 4       | 4         | 83     | 1            |
| 89     | 1            | 11         | 7        | 4       | 4         | 95     | 1            |
| 90     | 1            | 10         | 4        | 4       | 4         | 83     | 1            |
| ...    | ...          | ...        | ...      | ...     | ...       | ...    | ...          |
| 363    | 3            | 36         | 36       | 9       | 8         | 268    | 1            |
| 364    | 2            | 18         | 5        | 5       | 2         | 66     | 0            |
| 365    | 1            | 1          | 0        | 0       | 0         | 3      | 0            |

Столбцы таблицы 2 обозначают: `userid` – идентификатор студента, `course_tried` – количество начатых курсов, `study_days` – количество дней обучения, `reviewed` – количество просмотров результатов, `started` – количество начатых заданий, `submitted` – количество завершенных заданий, `viewed` – количество просмотров модулей курсов, `stay_studied` – учится/отчислен после первого семестра.

Второй этап – обучение модели. Для обучения модели используется класс `DecisionTreeClassifier` библиотеки `Scikit-learn`. Подбор оптимальных значений параметров дерева осуществляется с помощью класса `GridSearchCV` – поиск наилучшего набора параметров, доставляющих минимум ошибке перекрестного контроля (`cross-validation`). В итоге оптимальное дерево решений выглядит следующим образом.

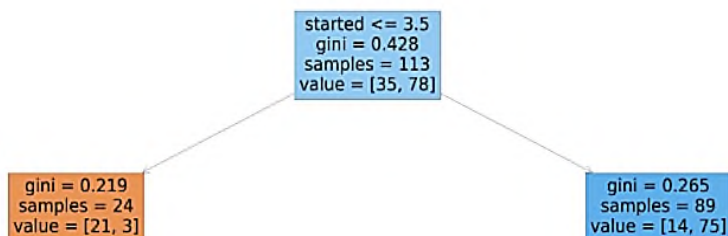


Рис. 1 Оптимальное дерево решений

Третий этап – прогнозирование и оценка качества модели. Прогнозная выборка – набор данных о студентах, зачисленных в 2019 году. Для оценки качества бинарной классификации проводится

ROC-анализ. В данном случае значение ROC = 0,64, что говорит о удовлетворительном качестве классификатора.

Следующим этапом должна стать отладка модели и ее внедрение в образовательную практику.

### **Выводы**

1. Образовательная аналитика – перспективное и быстро развивающееся научное направление, результаты внедрения которого могут существенно изменить подходы к управлению учебным процессом.

2. Разработка моделей для интеллектуального анализа образовательных данных является актуальной и важной задачей.

3. Для быстрого построения систем интеллектуального анализа данных может быть весьма эффективно использован язык программирования Python.

### **Библиография**

1. National learning platforms and tools [Электронный ресурс] // Web-site UNESCO. – URL: <https://en.unesco.org/covid19/educationresponse/nationalresponses> (Дата обращения: 23.08.2020).

2. By the Numbers: MOOCs During the Pandemic [Электронный ресурс] – URL: <https://www.classcentral.com/report/mooc-stats-pandemic> (Дата обращения: 23.08.2020).

3. Эпидемия коронавируса: воздействие на сферу образования [Электронный ресурс] / Департамент международного и регионального сотрудничества СП РФ. – Москва, 2020. – Режим доступа: <https://ach.gov.ru/upload/pdf/Covid-19-edu.pdf> (Дата обращения: 23.08.2020).

4. Online Education Statistics [Электронный ресурс] – URL: <https://educationdata.org/online-education-statistics/> (Дата обращения: 24.08.2020).

5. Jordan, K. MOOC completion rates: The data. [Электронный ресурс] / K. Jordan, 2014 – URL: [https://www.researchgate.net/publication/249994962\\_MOOC\\_completion\\_rates\\_The\\_data](https://www.researchgate.net/publication/249994962_MOOC_completion_rates_The_data) (Дата обращения: 24.08.2020).

6. Long, P., Siemens, G. Penetrating the Fog: Analytics in Learning and Education. [Электронный ресурс]: <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education> (Дата обращения: 24.08.2020).

7. Romero, C. Data mining in education / C. Romero, S. Ventura // Wiley interdisciplinary reviews. Data mining and knowledge discovery. 2013. – 3(1). – P. 12–27.

8. Siemens, G., Dawson, S., Lynch, G. Improving the Quality and Productivity of the Higher Education Sector [Электронный ресурс] – URL: [https://solarresearch.org/wp-content/uploads/2017/06/SoLAR\\_Report\\_2014.pdf](https://solarresearch.org/wp-content/uploads/2017/06/SoLAR_Report_2014.pdf) (Дата обращения: 24.08.2020).

9. Larusson, J.A., White, B. Learning Analytics: From Research to Practice / J.A. Larusson, B. White. New York: Springer Science+Business Media, 2014 – 191 p.

10. Kaggle ML & DS Survey [Электронный ресурс] URL: <https://www.kaggle.com/kaggle/kaggle-survey-2018> (Дата обращения: 23.08.2020).