

## Интеллектуальный анализ образовательных данных учебной дисциплины с использованием программы Orange

Китаева Ольга Игоревна

Иркутский национальный исследовательский технический университет,

Россия, Иркутск, *koi\_fk@mail.ru*

**Аннотация.** В статье рассмотрено понятие интеллектуального анализа образовательных данных и выполнен анализ на основе данных учебной дисциплины ВУЗа. Актуальность темы обусловлена необходимостью повышения качества образовательного процесса. Основным методом является получение интеллектуально улучшенных визуальных форм данных для более эффективного понимания скрытых в них закономерностей. В работе рассмотрены возможности применения программного продукта с открытым кодом «Orange» для реализации интеллектуального анализа данных текущей и промежуточной аттестаций студентов-бакалавров по учебной дисциплине «Исследование операций». В результате проведенного анализа выявлены проблемы в процессе освоения дисциплины и предложены пути их решения.

**Ключевые слова:** интеллектуальный анализ образовательных данных, Orange, анализ текущей успеваемости студентов, визуализация

**Цитирование:** Китаева О.И. Интеллектуальный анализ образовательных данных учебной дисциплины с использованием программы Orange / О.И. Китаева // Информационные и математические технологии в науке и управлении. – 2023. – № 1(29). – С. 190-200. – DOI:10.38028/ESI.2023.29.1.016.

**Введение.** Анализ данных используется в настоящее время для решения многих задач из разных областей. С его помощью можно исследовать, фильтровать, преобразовывать и моделировать данные для извлечения полезной информации и принятия решений. Интеллектуальный анализ образовательных данных (ИАОД) (от англ. Educational Data Mining, далее – EDM) – совокупность методов и алгоритмов анализа данных, накапливаемых в учебном заведении в процессе его деятельности с целью выявления скрытых, неочевидных, практически полезных и интерпретируемых знаний об учебном процессе и его участниках для поддержки и принятия решений [1]. Это направление исследований, связанное с применением методов интеллектуального анализа данных (Data mining) [2-4], машинного обучения и статистики к информации, производимой образовательными учреждениями. ИАОД направлен на исследование образовательных данных в целях улучшения качества образования [5].

На основе этого анализа возможно извлечь закономерности из данных, генерируемых в процессе обучения, таких как, например, посещаемость, успеваемость или данных из систем дистанционного обучения [6]. Процесс применения методов ИАОД, как правило, состоит из пяти следующих этапов:

1. Предварительная обработка данных.
2. Выявление закономерностей в данных тем или иным методом data mining.
3. Проверка обнаруженных закономерностей.
4. Применение найденных закономерностей для прогнозирования будущих событий в обучающей среде.
5. Использование построенных прогнозов при поддержке принятия решений и выработки образовательной политики.

Методы ИАОД позволяют обнаружить в данных знания об образовательном процессе и его участниках, которые были неизвестны до этого момента, и которые будут практически полезны и интерпретируемы для поддержки принятия решений [7-8]. Эти методы могут помочь в проектировании образовательной среды, в организации материалов учебных курсов и управлении образовательными ресурсами и дают возможность оценить влияние

используемых учебных стратегий на результаты успеваемости студентов. В работе «Прогнозирование результатов учебной деятельности студентов-первокурсников с помощью интеллектуального анализа образовательных данных» [9] авторы предлагают ряд типовых задач, которые могут быть решены средствами ИАОД. В частности, для обучающихся - осознанное формирование индивидуальной образовательной траектории, для преподавателей - оптимизация структуры и содержания лекционного курса, прогнозирование успешности учебной деятельности.

Различные аспекты работы образовательного учреждения могут быть проанализированы и использованы для поддержки принятия эффективных решений по управлению образовательным процессом. В том числе возможна схема образовательного процесса, когда детализированная информация о работе студента сохраняется в системе, а впоследствии используется для прогнозирования успешности обучающегося в будущем. [10].

В данном случае основной целью является преобразование имеющихся данных о процессе обучения в значимую информацию и поиск закономерностей для принятия решений, совершенствующих образовательные технологии.

**1. Оценка знаний студентов в системе электронного обучения Иркутского национального исследовательского технического университета (ИРНИТУ).** Очевидно, что качество знаний обучающихся зависит от многих факторов, в числе которых:

1. Качество организации учебного процесса (применение эффективных методик преподавания и современных информационных технологий);
2. Структура и содержание основных образовательных программ по направлениям подготовки;
3. Качество учебно-методического обеспечения;
4. Структура и содержание учебных планов дисциплин;
5. Мотивация студентов к обучению и другие.

Наряду с этим, качество знаний, получаемых обучающимися несомненно напрямую зависит от того, насколько последовательно и своевременно осваивается учебный материал.

Электронное обучение реализовано в ИРНИТУ на базе одной из наиболее востребованной и широко используемой системы MOODLE [11].

Эта система предоставляет преподавателям и студентам большой набор возможностей, в который входит:

1. Представление учебных материалов в структурированном виде;
2. Проведение тестирования;
3. Организация приема лабораторных и курсовых работ;
4. Обеспечение обратной связи со студентами (форум, чат, личные сообщения);
5. Отслеживание выполнения заданий студентами, журналы оценок и многие другие.

С помощью системы электронного обучения можно эффективно решать задачи коммуникативного взаимодействия между преподавателем и обучающимися, отслеживать:

- какие элементы курса были выполнены;
- фиксировать время выполнения и сдачи заданий;
- когда было сдано задание;
- насколько было просрочено время сдачи задания (если это имело место);
- оценку, выставленную преподавателем;
- комментарии преподавателя по работе, проделанной студентом.

К минусам стандартных отчетов MOODLE можно отнести отсутствие отчетов, демонстрирующих сопоставление действий пользователя на курсе (временной аспект) и

получаемые оценки за эти действия, что не позволяет быстро выявить связи между поведением студента на курсе и его успеваемостью, недостаточная наглядность и функциональность, чтобы сделать практически полезные выводы [11].

Вместе с тем, в системе MOODLE накапливается информация о результатах обучения и после соответствующей обработки возможно выявить поведенческие стратегии в обучении и проследить их связь с характеристиками обучающихся.

Рассмотренный в качестве примера электронный курс «Исследование операций», который изучают студенты-бакалавры в весеннем семестре второго курса обучения, предполагает выполнение курсовой и шести лабораторных работ. Электронный курс по этой дисциплине содержит как теоретическую составляющую, так и компоненты, позволяющие проводить тестирование по изученным разделам, проверять и оценивать отчеты по курсовой и лабораторным работам. Так компонент электронного курса «Отчет по лабораторной работе», содержит поля: Изображение пользователя, Фамилия Имя Отчество, Статус, Оценка, Последнее изменение (ответ), Отчет в виде файла, Комментарии к ответу, Последнее изменение (оценка), Итоговая оценка.

Компонент курса «Отчет по лабораторной работе» фактически представляет собой таблицу, в которой представлена вся информация, зафиксированная и оцененная в процессе выполнения лабораторной работы студентами конкретной группы. Просмотр данных предполагает переход от одной строки к другой и возможность анализировать, насколько своевременно выполнялась работа конкретным студентом, какие замечания он получал по отчетам, как была оценена его работа. Вместе с тем, изучая таблицу с данными, сложно получить информацию о динамике выполнения работы по всей группе студентов, обратить внимание на детали, пропустить важную информацию. Также интересно соотнести информацию текущего контроля с оценкой, полученной конкретным студентом во время промежуточной аттестации, и насколько своевременно был сдан экзамен. Исходя из этого, был определен набор данных (таблица 1), который был использован для анализа.

**Таблица 1.** Структура данных

Имя переменной	Назначение	Тип переменной	Роль переменной
index	номер студента в списке	числовой	характеристика
date1	дата сдачи отчета по лабораторной работе 1	дата	характеристика
grade1	оценка отчета по лабораторной работе 1	числовой	характеристика
date2	дата сдачи отчета по лабораторной работе 2	дата	характеристика
grade2	оценка отчета по лабораторной работе 2	числовой	характеристика
date3	дата сдачи отчета по лабораторной работе 3	дата	характеристика
grade3	оценка отчета по лабораторной работе 3	числовой	характеристика
date4	дата сдачи отчета по лабораторной работе 4	дата	характеристика
grade4	оценка отчета по лабораторной работе 4	числовой	характеристика
date5	дата сдачи отчета по лабораторной работе 5	дата	характеристика

grade5	оценка отчета по лабораторной работе 5	числовой	характеристика
date6	дата сдачи отчета по лабораторной работе 6	дата	характеристика
grade6	оценка отчета по лабораторной работе 6	числовой	характеристика
exam	экзаменационная оценка	числовой	характеристика
option	своевременность сдачи экзамена	категориальный	цель

**2. Программный продукт Orange Data Mining.** Аналитическая система Orange – это программа с открытым исходным кодом для машинного обучения и визуализации данных, обладающая большим набором исследовательских функций.

Программный продукт Orange (рус. Оранж), разработан Лабораторией биоинформатики Люблянского университета, предназначен для интеллектуального анализа данных (ИАД), статистических исследований и визуализации данных. Компоненты аналитической платформы называются виджетами, и варьируются от минималистичной визуализации данных, выбора подмножеств и предварительной обработки до эмпирической оценки алгоритмов обучения и прогностического моделирования [12-14].

В программном обеспечении Orange Data Mining применяется визуальное программирование, которое реализуется удобным графическим интерфейсом. В рамках визуального программирования аналитические процедуры создаются путём связывания предопределённых или разработанных пользователем блоков (виджетов), в то время как, более квалифицированные пользователи могут использовать Orange в качестве программной библиотеки Python для манипулирования данными и создания новых блоков (виджетов).

Структурно Orange состоит из интерфейса Canvas, на который пользователь помещает виджеты и создает рабочий процесс анализа данных. Виджеты предлагают базовые функции, такие как чтение данных, отображение таблицы данных, выбор функций, предикторы обучения, сравнение алгоритмов обучения, визуализация элементов данных и т. д. Пользователь может интерактивно исследовать визуализации или передавать выбранное подмножество в другие виджеты.

**3. Анализ образовательных данных средствами Orange Data Mining.** Средствами Orange Data Mining построена схема так называемого разведочного анализа данных, целью которого является, как правило, анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий. Схема приведена на рисунке 1.

В данном случае здесь задействованы Компоненты **Data Table** (Таблица данных), **Scatter Plot** (Точечный график), **Distributions** (Распределения), **Feature Statistics** (Статистика функций).

Разведочный анализ данных позволяет опровергнуть или подтвердить наши предположения относительно данных, выдвинуть новые гипотезы относительно того, что может произойти, если понимать, какие переменные находятся под контролем и какими рычагами можно воспользоваться для улучшения показателей, например, число экзаменов, сданных позже установленного срока. В проведенном исследовании в первую очередь рассматривается время выполнения работ и его влияние на итоговые результаты освоения дисциплины. Представляет интерес на первый взгляд, очевидная гипотеза: студенты, умеющие организовать свое время и, как следствие, выполняющие все задания вовремя, имеют более высокие конечные результаты. Далее дано пояснение схемы, использованной для анализа (рис. 1) и приведены полученные результаты. Следует отметить, что

предварительная обработка данных в приведенной схеме не использована, т.к. в ней не было необходимости.

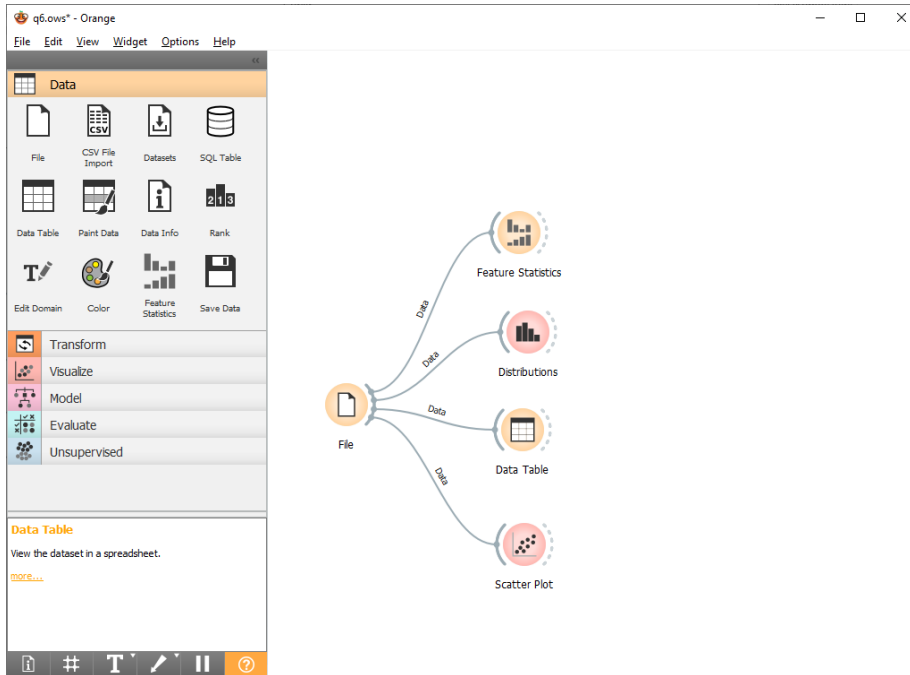


Рис. 1. Схема анализа данных в Orange Data Mining

Таблица данных (Data Table) – виджет принимает набор данных из входного файла и представляет их в виде таблицы. На рис. 2 показаны результаты использования компонента **Data Table** к анализируемым данным, структура которых показана в таблице 1.

option	index	date1	grade1	date2	grade2	date3	grade3	date4	grade4	date5	grade5	date6	grade6	exam
1	2022-03-23 00:00...	1	4	2022-04-06 00:00...	4	2022-04-19 00:00...	4	2022-05-01 00:00...	5	2022-05-22 00:00...	5	2022-05-25 00:00...	5	4
2	2022-06-04 00:00...	2	4	2022-06-02 00:00...	4	2022-06-01 00:00...	4	2022-06-01 00:00...	3	2022-06-01 00:00...	5	2022-06-02 00:00...	4	3
3	2022-05-27 00:00...	3	4	2022-05-29 00:00...	3	2022-05-31 00:00...	4	2022-05-27 00:00...	4	2022-05-27 00:00...	5	2022-05-11 00:00...	3	5
4	2022-05-18 00:00...	4	5	2022-04-17 00:00...	5	2022-04-25 00:00...	5	2022-05-13 00:00...	5	2022-05-15 00:00...	4	2022-05-20 00:00...	5	5
5	2022-03-30 00:00...	5	5	2022-03-27 00:00...	5	2022-03-27 00:00...	5	2022-05-07 00:00...	5	2022-05-07 00:00...	5	2022-05-23 00:00...	5	5
6	2022-04-03 00:00...	6	4	2022-04-14 00:00...	4	2022-05-11 00:00...	5	2022-05-07 00:00...	5	2022-05-08 00:00...	5	2022-05-13 00:00...	4	5
7	2022-03-18 00:00...	7	5	2022-05-13 00:00...	5	2022-05-11 00:00...	5	2022-05-01 00:00...	5	2022-05-02 00:00...	5	2022-05-13 00:00...	5	4
8	2022-04-15 00:00...	8	5	2022-04-22 00:00...	5	2022-05-17 00:00...	5	2022-05-20 00:00...	5	2022-05-27 00:00...	5	2022-06-02 00:00...	5	5
9	2022-03-18 00:00...	9	5	2022-04-01 00:00...	5	2022-04-08 00:00...	5	2022-05-21 00:00...	5	2022-05-05 00:00...	5	2022-05-18 00:00...	5	5
10	2022-04-20 00:00...	10	5	2022-05-04 00:00...	5	2022-05-06 00:00...	5	2022-05-11 00:00...	5	2022-05-16 00:00...	4	2022-05-22 00:00...	5	4
11	2022-03-11 00:00...	11	5	2022-04-01 00:00...	5	2022-04-01 00:00...	5	2022-04-01 00:00...	5	2022-04-28 00:00...	5	2022-05-13 00:00...	5	5
12	2022-03-14 00:00...	12	5	2022-03-31 00:00...	5	2022-04-13 00:00...	5	2022-04-15 00:00...	5	2022-05-15 00:00...	5	2022-05-18 00:00...	5	5
13	2022-04-12 00:00...	13	5	2022-05-19 00:00...	5	2022-05-11 00:00...	5	2022-05-26 00:00...	5	2022-06-02 00:00...	3	2022-06-03 00:00...	5	5
14	2022-04-27 00:00...	14	4	2022-05-01 00:00...	4	2022-05-03 00:00...	4	2022-05-11 00:00...	4	2022-05-11 00:00...	4	2022-05-16 00:00...	5	4
15	2022-11-20 00:00...	15	4	2022-11-21 00:00...	4	2022-11-21 00:00...	7	2022-11-21 00:00...	7	2022-11-18 00:00...	7	2022-11-21 00:00...	7	7
16	2022-05-26 00:00...	16	4	2022-05-26 00:00...	4	2022-05-29 00:00...	4	2022-05-29 00:00...	4	2022-05-29 00:00...	5	2022-05-29 00:00...	5	3
17	2022-05-29 00:00...	17	4	2022-05-24 00:00...	5	2022-05-29 00:00...	5	2022-05-29 00:00...	5	2022-05-29 00:00...	4	2022-05-29 00:00...	5	4
18	2022-03-29 00:00...	18	5	2022-04-10 00:00...	5	2022-05-11 00:00...	5	2022-05-01 00:00...	5	2022-05-10 00:00...	5	2022-05-13 00:00...	5	5
19	2022-04-28 00:00...	19	4	2022-04-27 00:00...	4	2022-05-11 00:00...	5	2022-05-11 00:00...	5	2022-05-04 00:00...	5	2022-05-25 00:00...	5	4
20	2022-03-23 00:00...	20	4	2022-04-30 00:00...	4	2022-05-04 00:00...	5	2022-05-03 00:00...	5	2022-05-04 00:00...	5	2022-05-18 00:00...	5	5
21	2022-05-17 00:00...	21	4	2022-05-21 00:00...	4	2022-05-26 00:00...	4	2022-05-23 00:00...	4	2022-05-23 00:00...	5	2022-05-29 00:00...	5	5
22	2022-04-15 00:00...	22	4	2022-04-21 00:00...	5	2022-05-25 00:00...	5	2022-05-11 00:00...	4	2022-05-21 00:00...	5	2022-05-21 00:00...	5	4
23	2022-03-29 00:00...	23	4	2022-04-23 00:00...	5	2022-04-23 00:00...	5	2022-05-15 00:00...	5	2022-05-15 00:00...	5	2022-05-22 00:00...	5	5
24	2022-04-19 00:00...	24	5	2022-04-18 00:00...	5	2022-04-18 00:00...	5	2022-05-11 00:00...	5	2022-05-11 00:00...	5	2022-05-08 00:00...	5	5
25	2022-04-25 00:00...	25	4	2022-04-25 00:00...	5	2022-05-11 00:00...	5	2022-05-01 00:00...	5	2022-05-16 00:00...	5	2022-05-16 00:00...	5	4
26	2022-04-05 00:00...	26	7	2022-04-05 00:00...	7	2022-04-06 00:00...	7	2022-04-22 00:00...	7	2022-04-20 00:00...	7	2022-05-16 00:00...	7	5
27	2022-10-18 00:00...	27	7	2022-10-18 00:00...	7	2022-10-18 00:00...	7	2022-10-18 00:00...	7	2022-10-18 00:00...	7	2022-10-18 00:00...	7	3
28	2022-05-23 00:00...	28	7	2022-06-01 00:00...	7	2022-06-01 00:00...	7	2022-06-01 00:00...	7	2022-06-01 00:00...	7	2022-06-04 00:00...	7	4
29	2022-05-23 00:00...	29	7	2022-05-23 00:00...	7	2022-05-23 00:00...	7	2022-05-26 00:00...	7	2022-05-26 00:00...	7	2022-05-25 00:00...	7	5
30	2022-05-25 00:00...	30	7	2022-05-23 00:00...	7	2022-05-23 00:00...	7	2022-05-26 00:00...	7	2022-05-26 00:00...	7	2022-05-25 00:00...	7	4
31	2022-04-12 00:00...	31	7	2022-04-19 00:00...	7	2022-05-16 00:00...	7	2022-04-20 00:00...	4	2022-04-28 00:00...	7	2022-05-23 00:00...	7	5
32	2022-05-03 00:00...	32	7	2022-05-03 00:00...	7	2022-05-13 00:00...	7	2022-05-12 00:00...	7	2022-05-12 00:00...	7	2022-06-02 00:00...	7	4
33	2022-03-22 00:00...	33	7	2022-04-01 00:00...	7	2022-04-19 00:00...	7	2022-04-19 00:00...	7	2022-04-28 00:00...	7	2022-05-05 00:00...	7	5
34	2022-03-21 00:00...	34	7	2022-04-27 00:00...	7	2022-04-28 00:00...	7	2022-05-19 00:00...	7	2022-05-20 00:00...	7	2022-06-01 00:00...	7	5
35	2022-05-27 00:00...	35	7	2022-05-03 00:00...	7	2022-04-01 00:00...	7	2022-05-17 00:00...	7	2022-06-01 00:00...	7	2022-06-01 00:00...	7	4
36	2022-04-05 00:00...	36	7	2022-05-17 00:00...	7	2022-05-26 00:00...	7	2022-05-27 00:00...	7	2022-05-23 00:00...	7	2022-05-26 00:00...	7	5
37	2022-10-24 00:00...	37	7	2022-10-24 00:00...	7	2022-10-24 00:00...	7	2022-10-24 00:00...	7	2022-10-24 00:00...	7	2022-10-24 00:00...	7	7
38	2022-05-01 00:00...	38	7	2022-05-07 00:00...	7	2022-06-02 00:00...	7	2022-05-07 00:00...	7	2022-05-07 00:00...	7	2022-05-24 00:00...	7	4
39	2022-03-14 00:00...	39	7	2022-03-31 00:00...	7	2022-04-13 00:00...	7	2022-04-15 00:00...	7	2022-05-03 00:00...	7	2022-05-13 00:00...	7	5
40	2022-04-21 00:00...	40	7	2022-04-07 00:00...	7	2022-04-13 00:00...	7	2022-04-18 00:00...	7	2022-05-06 00:00...	7	2022-05-06 00:00...	7	5
41	2022-04-27 00:00...	41	7	2022-05-01 00:00...	7	2022-05-03 00:00...	7	2022-05-11 00:00...	7	2022-05-11 00:00...	7	2022-05-16 00:00...	7	4
42	2022-04-05 00:00...	42	7	2022-04-19 00:00...	7	2022-04-29 00:00...	7	2022-04-29 00:00...	7	2022-05-13 00:00...	7	2022-05-13 00:00...	7	5
43	2022-03-07 00:00...	43	7	2022-03-30 00:00...	7	2022-04-06 00:00...	7	2022-04-13 00:00...	7	2022-04-10 00:00...	7	2022-05-15 00:00...	7	5
44	2022-04-04 00:00...	44	7	2022-05-26 00:00...	7	2022-05-26 00:00...	7	2022-05-26 00:00...	7	2022-05-27 00:00...	7	2022-05-27 00:00...	7	4
45	2022-04-03 00:00...	45	7	2022-04-29 00:00...	7	2022-05-09 00:00...	7	2022-05-12 00:00...	7	2022-05-17 00:00...	7	2022-05-17 00:00...	7	5

Рис. 2. Результаты вывода таблицы данных (виджет **Data Table**)

В левой части таблицы указано число анализируемых записей – 50. Т.к. переменная **option** (своевременность сдачи экзамена) определена как цель, все записи в таблице, соответствующие студентам, сдавшим экзамен не вовремя, подчеркнуты голубой чертой, записи студентов, сдавших экзамен вовремя, подчеркнуты красной. Просматривая таблицу, можно анализировать, как динамика выполнения работ повлияла на конечный результат. На рисунках 3-7 показаны результаты применения компонента **Distributions** для построения

диаграмм, демонстрирующих динамику сдачи отчетов по лабораторным работам 1-6 соответственно, единица измерения на графике одна неделя (1 week).

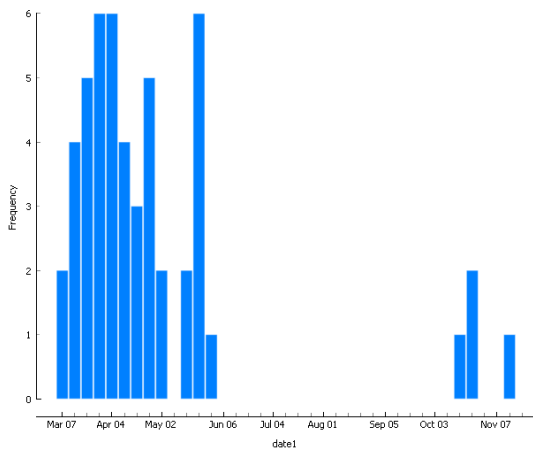


Рис. 3. Диаграмма времени сдачи отчетов по лабораторной работе 1

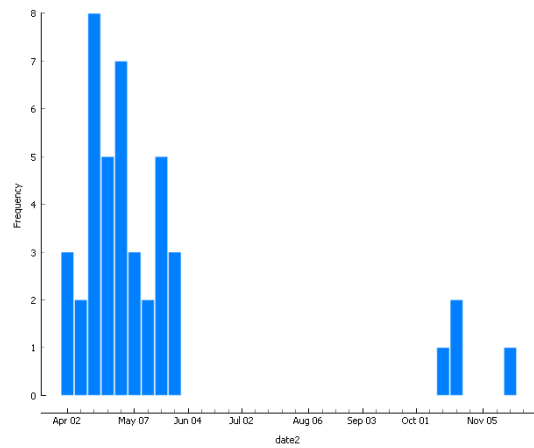


Рис. 4. Диаграмма времени сдачи отчетов по лабораторной работе 2

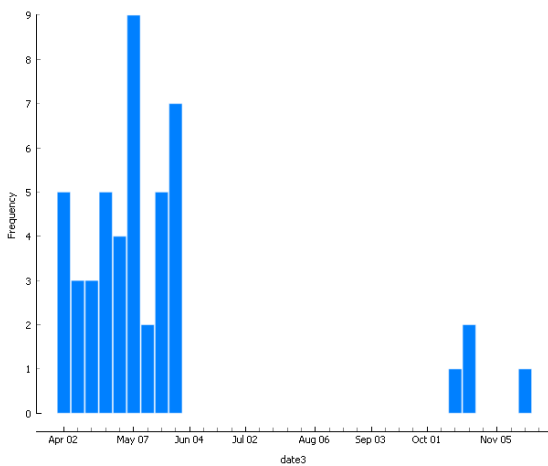


Рис. 5. Диаграмма времени сдачи отчетов по лабораторной работе 3

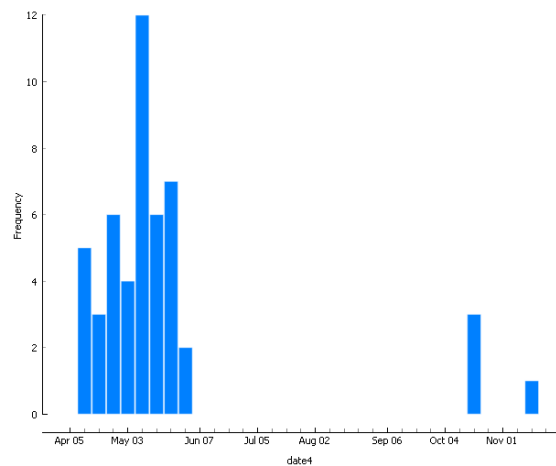


Рис. 6. Диаграмма времени сдачи отчетов по лабораторной работе 4

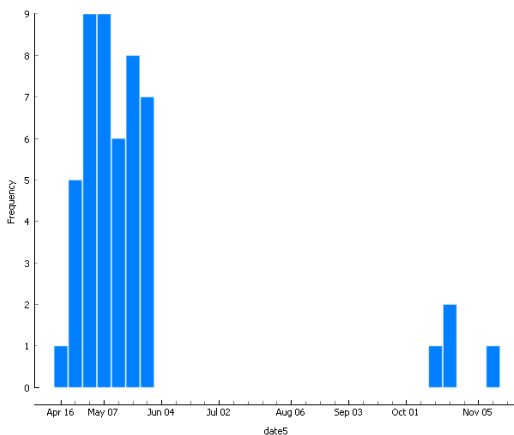


Рис. 7. Диаграмма времени сдачи отчетов по лабораторной работе 5

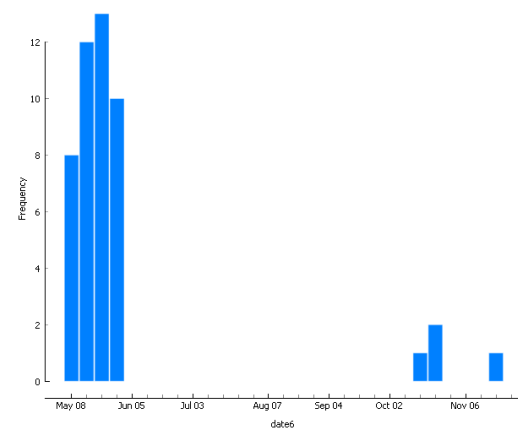
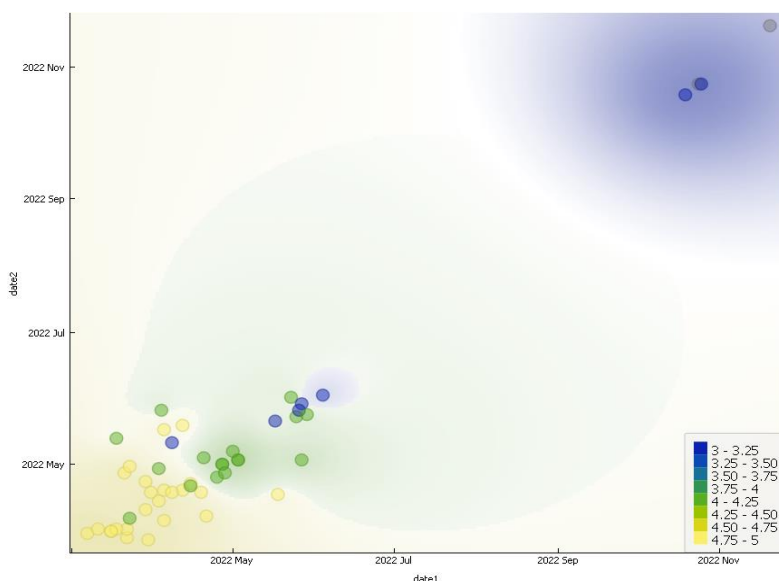


Рис. 8. Диаграмма времени сдачи отчетов по лабораторной работе 6

**Scatter Plot** выполняет визуализацию точечных графиков с предварительным анализом и интеллектуальными улучшениями визуализации данных. Очевидно, что основная цель – создать визуальную форму для лучшего и более эффективного понимания закономерностей, скрытых в данных.

Результаты построения точечного графика – применение виджета **Scatter Plot** – представлены на рис. 9. Как видно из рисунка, здесь фактически использованы три измерения: оси x, y и цвет. По оси x показана дата сдачи отчета по первой лабораторной работе **date1**, по оси y - дата сдачи отчета по второй лабораторной работе **date2**, цветом выделена оценка, полученная на экзамене. Рисунок демонстрирует вполне очевидный результат – студенты, раньше всех сдавшие отчеты, получили в основном отличные оценки.



**Рис. 9.** Результат применения компонента **Scatter Plot**

Следующий из использованных виджетов – **Feature Statistics** дает информацию о текущем размере набора данных, количестве и типах функций. Фрагмент полученных результатов представлен на рис. 10. Измерение «цвет» в данном случае демонстрирует своевременность сдачи экзамена (переменная *option*). Аналогично выводу таблицы данных, красный цвет соответствует студентам, сдавшим экзамен вовремя, голубой – тем, кто сдал экзамен не вовремя. Таблица справа содержит статистические данные о каждой функции в наборе данных, т.к. на рисунке показаны характеристики переменных, имеющих тип дата, соответствующие значения имеют аналогичный тип. Последний столбец – количество пропущенных значений в данных, в приведенном примере нулевой.

**4. Использование интеллектуального анализа данных для улучшения результатов текущей аттестации студентов.** Полученные результаты позволяют визуально наблюдать текущую ситуацию работы студентов при изучении учебной дисциплины. Предлагается следующая последовательность действий: преподаватель по окончании учебного семестра собирает все данные хода изучения дисциплины: посещаемость студентов, сроки выполнения запланированных работ, оценки качества выполнения работ и др. и выполняет их интеллектуальный анализ. По результатам анализа рекомендуются изменения, которые необходимо привнести для качественного улучшения всех рассматриваемых характеристик процесса обучения. Возможные изменения показаны в терминах системы менеджмента качества, как совокупности взаимосвязанных и взаимодействующих элементов для руководства и управления процессом обучения.

Предлагаемая корректировка процесса изучения учебной дисциплины «Исследование операций» приведена в таблице 2. Источником указанных в таблице 2 несоответствий являются результаты проведенного исследования (рис. 3-8).

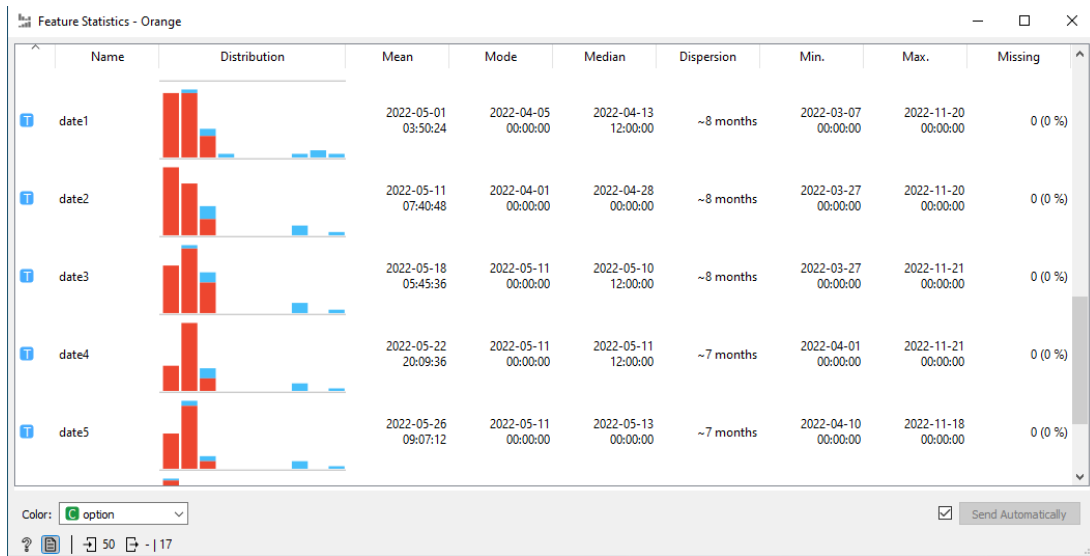


Рис. 10. Результат применения компоненты Feature Statistics

Таблица 2. Корректировка процесса изучения учебной дисциплины

Выявленные несоответствия	Корректирующие действия	Исполнитель корректирующих действий
Отчеты по лабораторным работам выполняются и сдаются существенно позже их выполнения	Установить и регистрировать в системе сроки сдачи работ	Преподаватель, ответственный за актуальность электронного курса в MOODLE
Большинство студентов сдают отчеты по лабораторным работам в конце учебного семестра	Определить контрольные точки сдачи работ в системе MOODLE	Преподаватель, ответственный за актуальность электронного курса в MOODLE
В действующей рабочей программе дисциплины (РПД) предполагается изучение разделов, на основании которых выполняются лабораторные работы фактически до конца учебного семестра.	Изменить последовательность изучения разделов дисциплины.	Составитель РПД
Отдельные студенты выполняют и сдают работы с существенным опозданием	Выявлять студентов, не выполняющих работы вовремя, и выяснять причины отставания	Преподаватель, ведущий лабораторные занятия
Студенты не умеют рационально использовать свое учебное время, имеют неравномерный темп обучения	Объяснять студентам необходимость повышения уровня самодисциплины	Дирекция института, кураторы



**Заключение.** Интеллектуальный анализ образовательных данных позволяет использовать данные об учебном процессе и его участниках для выявления скрытых, неочевидных, практически полезных и интерпретируемых знаний. Научная новизна предлагаемого подхода состоит в обосновании применения ИАОД, с использованием программы Orange Data Mining, для анализа организации учебного процесса с целью повышения его качества и рекомендаций по использованию интеллектуального анализа данных на основе выполненного исследования.

В ходе исследования была собрана схема разведочного анализа данных электронного курса «Исследование операций», изучаемого студентами-бакалаврами. Использован программный продукт с открытым исходным кодом Orange data mining. Визуальные данные, полученные программой Orange, указали на ряд негативных моментов, связанных как с текущей, так и промежуточной аттестацией студентов. В результате предложены корректирующие действия для улучшения качества образовательного процесса и, как следствие, повышения уровня текущей и промежуточной аттестации студентов.

### Список источников

1. Оськин А.Ф. Применение интеллектуального анализа образовательных данных для прогнозирования успешности учебной деятельности / А.Ф. Оськин, Д.А. Оськин // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки, 2016. – № 4. – URL: <https://elib.psu.by/bitstream/123456789/17089/5/8-12.pdf> (дата обращения: 29.01.2023).
2. Дюк В. Data Mining / В. Дюк, А. Самойленко. – Издательство: Питер. Серия: Учебный курс, 2001. – 368 с.
3. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям / Н.Б. Паклин, В.И. Орешков. – СПб.: Изд.: Питер, – 2013. – 704 с.
4. Агаев Ф.Т. Прогнозирование успеваемости студентов в электронном образовании с использованием методов data mining / Ф.Т. Агаев, Г.А. Мамедова, Р.Т. Меликова. – URL: <https://bik.sfu-kras.ru/ft/LIB2/ELIB/b74/free/i-489634964.pdf>(дата обращения: 27.01.2023).
5. Трубоч Г.Г. Использование анализа данных в образовании / Г.Г. Трубоч // 77-я научная конференция студентов и аспирантов Белорусского государственного университета. – Минск, 2020. – URL: <https://www.elibrary.ru/item.asp?id=46204873> (дата обращения: 29.01.2023).
6. Белоножко П.П. Анализ образовательных данных: направления и перспективы применения / П.П. Белоножко, А.П. Карпенко, Д.А. Храмов // Интернет-журнал НАУКОВЕДЕНИЕ, 2017. – Том 9. – № 4. – URL: <http://naukovedenie.ru/PDF/15TVN417.pdf> (дата обращения: 29.01.2023).
7. Горлушкина Н.Н. Задачи и методы интеллектуального анализа образовательных данных для поддержки принятия решений / Н.Н. Горлушкина, И.Ю. Коцюба, М.В. Хлопотов // Образовательные технологии и общество, 2015. – Т. 18. – № 1. – С. 472-482. – URL: <https://cyberleninka.ru/article/n/zadachi-i-metody-intellektualnogo-analiza-obrazova> (дата обращения: 29.01.2023).
8. Зорина Н.В. Интеллектуальный анализ образовательных данных: ретроспектива и перспективы развития / Н.В. Зорина, В.М. Панченко, А.Л. Зорин // Научная дискуссия: вопросы математики, физики, химии, биологии, 2017. – М., Изд. «Интернаука». – № 5-6(38). – С. 42-52.
9. Оськин А.Ф., Оськин Д.А., Рачицкий О.И. Прогнозирование результатов учебной деятельности студентов-первокурсников с помощью интеллектуального анализа образовательных данных – URL: <https://elib.psu.by/bitstream/123456789/21611/1/%d0%9e%d1%81%d1%8c%d0%ba %d0%b8%d0%bd-26-28.pdf> (дата обращения: 21.01.2023).
10. Есин Т.Е. Глухих. И.Н. Использование методов интеллектуального анализа данных для получения знаний об образовательном процессе. – URL: <https://core.ac.uk/download/pdf/226255621.pdf> (дата обращения: 23.01.2023).
11. Тербушева Е.А. Аналитический потенциал платформы Moodle для мониторинга качества персонализированного обучения / Е.А. Тербушева, К.Р. Пиотровская. // Общество Коммуникация. Образование, 2021. – Т. 12. – № 4. – С. 19-34. – DOI: 10.18721/JHSS.12402
12. Описание системы Orange. – URL: <http://soware.ru:3000/products/orange> (дата обращения: 29.01.2023).
13. Юсупов Н. Исследование методов классификации в программе Orange / Н. Юсупов, А. Савельева, О.Г. Леонова // Молодежная школа-семинар по проблемам управления в технических системах имени А.А. Вавилова, 2020. – Т. 1. – С. 27-30.

14. Юсупов Н. Исследование методов кластеризации в программе Orange / Н. Юсупов. // Молодежная школа-семинар по проблемам управления в технических системах имени А.А. Вавилова, 2020. – Т. 1. – С. 35-37.

*Китаева Ольга Игоревна. Старший преподаватель института информационных технологий и анализа данных Иркутского национального исследовательского технического университета, руководитель отдела по работе со студентами, сфера научных интересов: Интеллектуальный анализ образовательных данных.*

UDC 378.14.015.62

DOI:10.38028/ESI.2023.29.1.016

## Intelligent analysis of educational data of the educational discipline using the Orange program

Olga I. Kitaeva

Irkutsk National Research Technical University,  
Russia, Irkutsk, *koi\_fk@mail.ru*

**Abstract.** The article considers the concept of intellectual analysis of educational data and analyzes the educational data of the academic discipline of the university. The relevance of the topic is due to the need to improve the quality of the educational process. The main method is to obtain intellectually improved visual forms of data for a more effective understanding of the patterns hidden in them. The paper considers the possibilities of using the open source software product "Orange" for the implementation of data mining for the current and intermediate certification of bachelor students in the academic discipline "Operations Research". As a result of the analysis, problems were identified in the process of mastering the discipline and ways to solve them were proposed.

**Keywords:** intellectual analysis of educational data, Orange, analysis of current student performance, visualization

### References

1. Oskin A.F., Oskin D.A. Primenenie intellektualnogo analiza obrazovatelnyh dannyh dlya prognozirovaniya uspehnosti uchebnoj deyatel'nosti [Application of intellectual analysis of educational data to predict the success of educational activities]. Vestnik Polotskogo gosudarstvennogo universiteta. Seriya S. Fundamental'nyye nauki [Bulletin of Polotsk State University. Series C. Fundamental sciences], 2016, no. 4, available at: <https://elib.psu.by/bitstream/123456789/17089/5/8-12.pdf> (accessed: 01/29/2023).
2. Dyuk V., Samojlenko A. Data Mining: uchebnyj kurs [Data Mining: training course], 2001, Izdatel'stvo: Piter. Seriya: Uchebnyj kurs [Publisher: Piter. Series: Training course], 368 p.).
3. Paklin N. B., Oreshkov V. I. Biznes-analitika: ot dannyh k znaniyam [Business Analytics: From Data to Knowledge] SPb., Piter [Piter], 2013, 704 p.
4. Agaev F. T., Mamedova G. A., Melikova R. T. Prognozirovanie uspevaemosti studentov v elektronnom obrazovanii s ispolzovaniem metodov data mining [Forecasting student performance in e-education using data mining methods]. Available at: <https://bik.sfu-kras.ru/ft/LIB2/ELIB/b74/free/i-489634964.pdf> (accessed: 01/27/2023).
5. Trubach G.G. Ispolzovanie analiza dannyh v obrazovanii [The use of data analysis in education, in the collection]: 77th scientific conference of students and graduate students of the Belarusian State University, 2020, available at: <https://www.elibrary.ru/item.asp?id=46204873> (accessed: 01/29/2023).
6. Belonozhko P.P., Karpenko A.P., Hramov D.A. Analiz obrazovatelnyh dannyh: napravleniya i perspektivy primeneniya [Analysis of educational data: directions and prospects of application]. Internet-zhurnal NAUKOVEDENIYE [Internet journal NAUKOVEDENIE], 2017, vol. 9, no. 4, available at: <http://naukovedenie.ru/PDF/15TVN417.pdf> (accessed: 29/01/2023).
7. Gorlushkina N.N., Kocyuba I.Yu., Hlopotov M.V. Zadachi i metody intellektualnogo analiza obrazovatelnyh dannyh dlya podderzhki prinyatiya reshenij. [Tasks and methods of educational data mining for decision support]. Obrazovatel'nyye tekhnologii i obshchestvo [Educational technologies and society], 2015, vol. 18, no. 1, pp. 472-482, available at: <https://cyberleninka.ru/article/n/zadachi-i-metody-intellektualnogo-analiza-obrazova> (accessed: 01/29/2023).
8. Zorina N.V. Intellektualnyj analiz obrazovatelnyh dannyh: retrospektiva i perspektivy razvitiya [Intellectual analysis of educational data: retrospective and development prospects. Nauchnaya diskussiya: voprosy

- matematiki, fiziki, khimii, biologii [Scientific discussion: questions of mathematics, physics, chemistry, biology], 2017, M., Internauka, no. 5-6(38), pp. 42-52.
9. Oskin A.F., Oskin D.A., Rachickij O.I. Prognozirovanie rezultatov uchebnoj deyatel'nosti studentov-pervokursnikov s pomosh'yu intellektual'nogo analiza obrazovatel'nykh dannykh. [Forecasting the results of educational activity of first-year students using the intellectual analysis of educational data], available at: <https://elib.psu.by/bitstream/123456789/21611/1/%d0%9e%d1%81%d1%8c%d0%ba%d0%b8%d0%bd-26-28.pdf> (accessed: 21/01/2023).
  10. Esin T.E., Gluhii I.N. Ispol'zovanie metodov intellektual'nogo analiza dannykh dlya polucheniya znaniy ob obrazovatel'nom processe. [Using data mining methods to gain knowledge about the educational process], available at: <https://core.ac.uk/download/pdf/226255621.pdf> (accessed 01/23/2023).
  11. Terbusheva E.A., Piotrovskaya K.R. Analiticheskij potentsial platformy Moodle dlya monitoringa kachestva personalizirovannogo obucheniya [Analytical potential of the Moodle platform for monitoring the quality of personalized learning]. Obshchestvo Kommunikatsiya. Obrazovaniye [Society Communication. Education], 2021, vol. 12, no. 4, pp. 19-34, DOI: 10.18721/JHSS.12402.
  12. Opisanie sistemy Orange [Description of the Orange system]. available at: <http://soware.ru:3000/products/orange> (accessed: 01/29/2023).
  13. Yusupov N., Saveleva A., Leonova O.G. Issledovanie metodov klassifikatsii v programme Orange [Study of classification methods in the Orange program]. Molodezhnaya shkola-seminar po problemam upravleniya v tekhnicheskikh sistemakh imeni A.A. Vavilova [Youth school-seminar on control problems in technical systems named after A.A. Vavilov], 2020, vol. 1, pp. 27-30.
  14. Yusupov N. Issledovanie metodov klasterizatsii v programme Orange [Investigation of clustering methods in the Orange program]. Molodezhnaya shkola-seminar po problemam upravleniya v tekhnicheskikh sistemakh imeni A.A. Vavilova [Youth school-seminar on control problems in technical systems named after A.A. Vavilov], 2020, vol. 1, pp. 35-37.

**Kitaeva Olga Igorevna.** Senior lecturer at the institute of information technology and data analysis of the Irkutsk National Research Technical University, head of the student relations department, area of expertise: educational data mining.

Статья поступила в редакцию 20.02.2023; одобрена после рецензирования 28.02.2023; принята к публикации 09.03.2023.

The article was submitted 02/20/2023; approved after reviewing 02/28/2023; accepted for publication 03/09/2023.