

А. Д. Черемухин

К ВОПРОСУ ПОСТРОЕНИЯ ТАКСОНОМИИ ЗАДАЧ ПРИ ОБУЧЕНИИ АНАЛИЗУ ДАННЫХ

Нижегородский государственный инженерно-экономический университет,
Нижний Новгород, Россия

Поступила в редакцию 18.06.2024 г.

Принята к публикации 27.10.2024 г.

doi: 10.5922/vestnikpsy-2025-1-11

123

Для цитирования: Черемухин А. Д. К вопросу построения таксономии задач при обучении анализу данных // Вестник Балтийского федерального университета им. И. Канта. Сер.: Филология, педагогика, психология. 2025. №1. С. 123–134. doi: 10.5922/vestnikpsy-2025-1-11.

Широкое применение технологий искусственного интеллекта, анализ данных, ставка на них как на основу развития экономики будущего существенно увеличивают спрос на данных специалистов. В связи с этим необходимо обратить внимание на методiku преподавания анализа данных в университетах. Цель статьи – разработать основы таксономии для подготовки практических заданий по анализу данных и апробировать ее. Автором рассмотрены основные таксономии учебных задач, обоснована их слабая применимость к области анализа данных из-за ее мультидисциплинарности и многоаспектности. Предложена авторская горизонтально-вертикальная таксономия учебных задач на основе трех основных процессов – мышление, коммуникация, деятельность; при этом сложность задач предлагается взять в качестве ключевого показателя, на котором основана вертикальная часть таксономии. На примере темы «линейная регрессия» представлены типовые задачи, характерные для соответствующих первичных и вторичных процессов, а также их смешения. Разработаны и показаны конкретные задания для студентов уровня магистратуры направления подготовки «Бизнес-информатика» в рамках дисциплины «Компьютерный анализ данных» с использованием языка R. Обоснована необходимость дальнейших исследований в данном направлении, поставлен ряд вопросов для продолжения работы.

Ключевые слова: таксономия, учебная задача, анализ данных, практическое занятие

Введение

Появление больших языковых моделей как открывающей технологии технологического пакета «Искусственный интеллект» стало катализатором значительного числа дискуссий о возможности перестройки на его основе отдельных отраслей экономики и всей экономической сферы в целом. В условиях обострения геополитической ситуации задача



скорейшего повышения эффективности хозяйствования в стране стала стратегической — и автоматически стратегическое значение получил вопрос о повышении количества и качества выпускаемых специалистов в области ИИ, подготавливаемых системой образования.

Еще в 2019 г. правительством страны была принята дорожная карта по развитию федерального проекта «Нейротехнологии и искусственный интеллект», одной из задач которой является «повышение уровня обеспечения российского рынка технологий ИИ квалифицированными кадрами и уровня информированности населения о возможных сферах использования ИИ» [1]. Данный курс был подтвержден в конце 2023 г. заместителем министра науки и высшего образования Андреем Омельчуком, который заявил на конференции *Artificial Intellect Journey*, что к 2030 г. вузам нужно увеличить выпуск специалистов в области ИИ более чем в 5 раз.

Количественная составляющая поставленных целей решается путем административного и нормативно-правового регулирования, в том числе и корректировкой цифр приема на различные направления подготовки. Вопрос подготовки качественных специалистов — более сложный, актуальный и глубокий.

Существующие программы обучения по направлениям анализа данных и искусственного интеллекта, реализуемые в том числе и ведущими вузами, являются практикоориентированными, разработанными на основе опыта крупных компаний, прошедшими эмпирическую проверку эффективности. Однако возможность масштабирования данных подходов в условиях отсутствия теоретического обоснования еще не подтверждена. На конец мая 2024 г. поиск по сайту научной электронной библиотеки eLIBRARY по запросам «преподавание анализа данных», «преподавание искусственного интеллекта» выдал менее 100 релевантных результатов по всей базе, а поиск по базе диссертаций ВАК позволил констатировать наличие лишь единичных работ, посвященных методике преподавания данных предметов. Это диссертации Е. А. Тербушевой «Методика обучения студентов с профильной вузовской подготовкой в области математики интеллектуальному анализу данных» [2], А. А. Салаховой «Методика обучения основам искусственного интеллекта и анализа данных в курсе информатики на уровне среднего общего образования» [3] и П. А. Меренковой «Вариативное обучение системам искусственного интеллекта в рамках учебного предмета “Информатика” основной школы» [4]. С учетом того что лишь одна из рассмотренных диссертаций охватывает уровень высшего образования, можно констатировать недостаточный уровень развития научной дискуссии в области методики преподавания дисциплин, связанных с анализом данных и искусственным интеллектом.

Впрочем, данные проблемы характерны и для мировой педагогической науки. Юрий Демченко с коллегами из университета Амстердама отмечает, что «в настоящее время не существует ни общепризнанных учебных программ профессионального образования в области науки о данных, ни общих программ обучения технологиям больших данных. Кроме того, не существует единого подхода к эффективному построению учебных программ по науке о данных профессионального уровня.



Университеты как в Европе, так и в США (и за их пределами) не предлагают достаточных возможностей для подготовки большого количества специалистов нового типа» [5, р. 610].

Усугубляется это в том числе и терминологическими несостыковками: наличие многочисленных определений понятий «большие данные», «анализ данных», «искусственный интеллект», «Data Science» затрудняет методическую работу из-за размытости объекта исследования. Для устранения этого мы в дальнейшем будем пользоваться следующими определениями¹:

– *искусственный интеллект* есть способность компьютера обучаться, принимать решения и выполнять действия, свойственные человеческому интеллекту;

– *машинное обучение* есть класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач;

– *глубокое обучение* есть создание многослойных нейронных сетей в областях, где требуется более продвинутый или быстрый анализ и традиционное машинное обучение не справляется;

– *большие данные* – это область работы с большим объемом часто неструктурированных данных;

– *наука об анализе данных* – наука о процессе наделения смыслом массивов данных, методах визуализации, принятии решений на основе этих данных.

В такой трактовке цель данной работы заключается в разработке методики составления заданий для практических занятий по анализу данных для студентов высшего образования, обучающихся по профилю «Анализ данных и машинное обучение» в рамках направления подготовки 09.03.01 «Информатика и вычислительная техника».

Материалы и методы исследования

Составление обоснованных и эффективных заданий для любой предметной области требует предварительного содержательного анализа ее компонент. В работе [6], например, отмечено, что выпускник в сфере бизнес-аналитики должен обладать:

– знаниями в области математики и статистики (учащиеся должны знать набор инструментов для проведения сводного, классификационного или прогнозного анализа (не)структурированных данных, уметь строить базовые модели для анализа и т. д.);

– навыками в области программирования (учащийся должен ознакомиться с современными языками программирования с открытым исходным кодом, изучить инструменты высокого уровня для визуализации данных и т. д.);

– существенным опытом в предметной области (учащиеся должны применять аналитический подход к решению бизнес-задач, специфичных для конкретной предметной области; понимать, как интерпретировать результаты анализа и передавать их неспециалисту) [6, р. 373].

¹ Представлены Натальей Носенко из Сбера на Летней цифровой школе Сбера – 2023.



Данная трактовка иллюстрирует междисциплинарную природу анализа данных как предметной области, что было подтверждено и рядом других исследований [7]. Однако, по мнению их авторов, вопрос о компонентах области анализа данных как специфической предметной области с точки зрения педагогики остается дискуссионным.

В исследовании [8] отмечается важность и необходимость коммуникации при решении задач анализа данных в организациях, предложено выделить в проектной команде отдельную соответствующую специальность. На необходимость развития у аналитиков коммуникационных навыков обращает внимание и профессиональный стандарт 06.042 «Специалист по большим данным» [9]. Его первая трудовая функция на уровне бакалавриата требует наличия умений «проводить презентации при консультировании заказчика, согласовании и утверждении требований к результатам аналитических работ с использованием технологий больших данных» и «проводить переговоры с целью выявления требований заказчика к результатам анализа, формировать и согласовывать требования к результатам аналитических работ с использованием технологий больших данных». Вторая трудовая функция требует проведения действий по «распределению ролей и состава аналитических работ между участниками группы для анализа больших данных».

Исходя из этого, мы считаем, что при рассмотрении анализа данных как объекта педагогики целесообразно учитывать наличие в нем трех основных составляющих: математической, программистской и коммуникационной. Именно сочетанием столь разнородных составляющих объясняется и сложность подготовки студентов в данном направлении. Для построения эффективной системы обучения необходимо преодолеть весь комплекс проблем, связанных с преподаванием математики [10] и программирования [11], органично интегрировать задания, формирующие коммуникационные компетенции, и сохранить их практико-ориентированную направленность. При этом, как отмечается в [6], «аналитическая индустрия имеет лидерство в подготовке аналитиков, большинство из которых проходят внутреннее обучение в конкретной организации... Однако этот подход вряд ли позволит подготовить аналитиков с разнообразными навыками. Это связано с тем, что обучение в отдельной организации слишком приспособлено к конкретным ее задачам» [6, р. 372], поэтому большой уровень направленности на решение конкретных кейсов может привести к формированию слишком узкоспециализированных работников.

При этом вполне логичной является разработка заданий с учетом тех навыков, которые требуются работодателями, и задач, предлагаемых ими на собеседованиях. Приведем примеры таких задач:

В файле `revo_ds_test_task.csv` собраны данные по повторным займам текущих клиентов компании. По имеющемуся набору данных необходимо построить модель, которая будет прогнозировать значение целевой переменной `'bad_flag'`. Подготовить небольшую презентацию (1–2 слайда), в которой аргументируется выбор модели и показан предполагаемый уровень дефолтности при различных уровнях одобрения. Расчеты желательно предоставить в виде `python`-скрипта / `ipython` `notebook` [12].



Таблица (task2.txt) содержит 11 столбцов чисел. Первые 10 столбцов – входные переменные (x_1, \dots, x_{10}), 11-й столбец – выходная прогнозируемая переменная y . Каждая строка в файле – это один обучающий пример для построения статистической модели для зависимости $y = y(x_1, x_2, \dots, x_{10})$. Конкретный вид связи между выходной и входными переменными не известен. Требуется провести разведочный анализ данных, выбрать критерий для точности модели и указать степень важности (информативности) каждой из 10 входных переменных с точки зрения точности модели. Что еще можно предложить, чтобы улучшить получившийся результат? [13].

В представленных задачах отчетливо фиксируется присутствие трех ранее выделенных компонент: математической, программистской и коммуникационной. Соответственно, при возникновении вопроса о построении образовательных задач и их систем в области анализа данных он влечет за собой вопрос о поиске подходящей таксономии.

Самой часто применяемой в педагогической практике в области компьютерных наук [14–17] является классическая таксономия задач Блума [18]. Как отмечено в [19], в 46 % публикаций, посвященных методике преподавания программирования, используется именно она. Однако процесс применения ее на практике оказался сложным даже для опытных преподавателей [20] в том числе из-за сложности отнесения задач к конкретной ступени. Как отмечается в [21], разные учащиеся могут решать одну задачу разными способами рассуждения, относящимися к разным уровням таксономии.

Модифицированная Андерсоном [22] таксономия Блума значительно чаще встречается как основа научных работ западных педагогов, и критики ее применения в программировании отмечают, например, неопределенность понятий «применение процесса» / «создание процесса» применительно к контексту программирования [23], неполный и недостаточный уровень таксономий [24], отсутствие вычислительных примеров, иллюстрирующих уровни и подуровни таксономии, и вообще слабую ее приспособленность к концепции программирования [20].

Таксономия Марцано [25] предусматривает наличие трех систем. В рамках когнитивной системы выделяются четыре иерархических уровня: обретение – понимание – анализ – применение знания. Кроме когнитивной системы, выделяется «Я-система», которая описывает отношение человека к мышлению, а также «система метапознания», описывающая отношение человека к непосредственно происходящему процессу обучения.

Таксономия Толлингеровой, представленная в [26], делит все учебные задачи на 5 категорий: предполагающие воспроизведение знаний, предполагающие простые мыслительные операции, предполагающие сложные мыслительные операции, предполагающие обобщение знаний и сочинение, предполагающие продуктивное мышление.

В настоящее время появляется большое количество разных таксономий, основанных на смешении таксономий разного типа – эмоциональных, когнитивных и т.д., примеры которых представлены в [27]; при этом все они сложноприменимы к построению системы задач анализа данных из-за специфики данной области.



Кроме того, все они обладают недостатком, который можно обозначить как «линейность иерархии». Как отмечается в [28], «ни одна из линейных таксономий не отвечает тем потребностям, для которых она была разработана», потому что «обучение чаще всего носит нелинейный, рекурсивный и эмерджентный характер» [28, р. 21]. Соответственно, для междисциплинарной и многокомпонентной деятельности, такой как анализ данных, разложение ее в линейную иерархию не может привести к построению эффективной и понятной таксономии.

Неисследованным является вопрос о влиянии больших языковых моделей на таксономию во всей сфере высшего образования в целом. Широкое использование ИИ-ассистентов для написания кода в отрасли программирования, например, делает задачи первого типа (на воспроизводство) бессмысленными с точки зрения студентов (что отрицательно сказывается на мотивации) и ставит вопрос об их необходимости в таксономии и в обучении в принципе. При этом в области математики такого вопроса не возникает, поскольку математические знания больших языковых моделей находятся на значительно более низком уровне.

Учитывая междисциплинарность и многокомпонентность анализа данных как деятельности, мы предлагаем применить горизонтальный подход к построению таксономии. Поддерживая тезис о том, что «учебную задачу можно считать решенной только в том случае, если произошли заранее заданные изменения у того, кто решает задачу» [26, с. 100], определим учебную задачу как создаваемое преподавателем задание, вызывающее у студента при корректном выполнении заранее заданные изменения. При таком подходе вертикальная часть таксономии будет соответствовать сложности задач.

Отмеченное выше наличие трех компонент в области анализа данных (математическая, программистская, коммуникационная) может быть наложено на модифицированную С. Б. Переслегиным [29] схему мыследеятельности. Она подразумевает наличие трех основных (мышление, коммуникация и деятельность) и трех вторичных (понимание, рефлексия и смещение / сотворение) процессов. Таким образом, три основных процесса сочетаются с тремя основными компонентами таксономии: математическая компонента — мышление, программистская — деятельность (в данном случае программы выступают инструментом деятельности, анализ можно проводить и путем простых расчетов, как это делалось до появления ЭВМ), коммуникационная — коммуникация. Это позволяет конструировать задачи основных классов и смешивать их между собой.

Сами образовательные задачи разрабатывались для курса «Компьютерный анализ данных» по направлению подготовки 38.04.05 «Бизнес-информатика» (программа «Анализ больших данных в экономике»). Дисциплина рассчитана на 72 часа работы, из которых 26 — аудиторных (10 часов — лекционные занятия, 16 — практические), все практические занятия выполняются с использованием языка R. Опишем разработанную систему задач по теме «Построение и проверка линейной регрессии».



Результаты

Применительно к теме «Построение и проверка линейной регрессии» была сформулирована таблица соответствия процессов, целей задачи и примерного их содержания.

Возможные задачи по теме «Построение и проверка линейной регрессии»

Процесс	Цель задачи	Описание задачи
Понимание	Проанализировать	Дан код, реализующий регрессию — нужно понять, какая модель строится, по каким переменным, какие результаты
Рефлексия	Оценить	Дано решение задачи с выводами — нужно проверить правильность и эффективность кода и выводов
Смещение	Сотворить	Нужно найти, что и как в наибольшей степени влияет на зависимую переменную
Мышление	Придумать	Дана задача в языке пользователей результатов модели — необходимо составить план решения задачи
Деятельность	Написать код	Постройте уравнение регрессии с помощью такой-то функции и выведите результаты
Коммуникация	Описать	Дан результат выполнения кода — надо сделать по нему выводы
Мышление с оттенком деятельности	Применить алгоритм	Необходимо на основе поставленной задачи определить конкретный алгоритм регрессии, который будет применен, пакеты и функции программы
Деятельность с оттенком мышления	Обучить модель	Построить саму модель регрессии
Деятельность с оттенком коммуникации	Определить и собрать данные	Собрать, очистить необходимые для обучения данные
Коммуникация с оттенком деятельности	Предложить решение	Дана задача в языке пользователей результатов модели — необходимо составить план решения задачи
Коммуникация с оттенком мышления	Объяснить	Даны результаты выводов по модели. Что это значит в контексте прикладной задачи?
Мышление с оттенком коммуникации	Обсуждение гипотез	Объяснить свой способ решения поставленной задачи

Комплект заданий в практике не состоит из всех 12 типов заданий. Поскольку это практическое задание — первое в теме, была выбрана следующая структура:



Задание 1, процесс — понимание, цель задачи — проанализировать.

Вам дан код, реализующий регрессию

```
library(tidyverse)
```

```
library(AER)
```

```
data("CASchools")
```

```
glimpse(CASchools)
```

```
model_1 <- lm(income~read, data=CASchools)
```

```
summary(model_1)
```

По нему вам необходимо определить и описать, по каким данным строится регрессия, что выступает в роли зависимых и независимых переменных.

Задание 2, процесс — понимание, цель задачи — проанализировать.

Вам даны результаты построения модели регрессии (рис.). Определите, по каким данным строилась регрессия, что фигурировало в качестве зависимой и независимой переменных. Сделайте максимально возможное количество выводов о значимости регрессии по представленным данным, если это возможно, сделайте содержательные выводы по уравнению регрессии.

```
Call:
lm(formula = testscore ~ average, data = ProgramEffectiveness)

Residuals:
    Min       1Q   Median       3Q      Max
-9.2996 -1.8472  0.1343  2.8248  5.9916

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   11.853     4.434   2.673  0.0120 *
average         3.235     1.407   2.299  0.0287 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.657 on 30 degrees of freedom
Multiple R-squared:  0.1498,    Adjusted R-squared:  0.1214
F-statistic: 5.284 on 1 and 30 DF,  p-value: 0.02866
```

Рис. Результаты построения модели регрессии как условие для задания 2

Задание 3, процесс — деятельность и понимание, цель задачи — написать код и проанализировать.

Вам дан датафрейм (<https://cloud.mail.ru/public/unrw/m6jaxidnv>) и шпаргалка по нему (<https://cloud.mail.ru/public/hNyN/qC3K4oabJ>).

Постройте с помощью функции `lm()` уравнение линейной регрессии и сделайте по нему соответствующие выводы, если зависимая переменная — общее количество произведенного зерна, независимая — средняя влажность воздуха.

Задание 4, процесс — рефлексия, цель задачи — оценить.

Вам дан датафрейм (<https://cloud.mail.ru/public/unrw/m6jaxidnv>) и шпаргалка по нему (<https://cloud.mail.ru/public/hNyN/qC3K4oabJ>).



Также вам дан код, строящий уравнение регрессии и выводы по нему. Исполните данный код, сделайте вывод о том, насколько полно и правильно сделаны выводы по результатам регрессии.

```
df <- Dataframe5
model <- lm(Cost ~ Machine_Time, data = df)
summary(model)
# Вывод: Коэффициент детерминации = 0.6193, значит, модель объясняет 6.193 % дисперсии данных, модель # не очень точна.
# Прибыль увеличивается на 1.670e-03 млн руб
# Значение переменной p-value = 2e-16 < 0.05, значит, коэффициент наклона статистически значим
# Между прибылью и общей площадью посевов существует зависимость
```

131

Задание 5, процесс — смещение, цель задачи — сотворить.

Загрузите из пакета AER данные датафрейма Equipment. Постройте по этой базе данных статистически значимое уравнение парное линейное регрессии, используя в качестве зависимой переменной показатель valueadded.

Обсуждение

Представленная система задач, как и общий подход, была апробирована при реализации курса «Компьютерный анализ данных». Однако при ее комплектовании возник ряд вопросов, основной из которых — как «собирать» задачи в комплекты? Можно предложить несколько способов:

- линейно — в порядке возрастания сложности по одному из процессов;

- «лесенкой» — в порядке возрастания сложности по процессу и его модификациям (например, задача на мышление — мышление с оттенком коммуникации — мышление с оттенком деятельности — и снова на мышление);

- концентрически — обходятся все возможные задачи (вспомогательный вопрос — какой тип задач должен быть первым) на определенном уровне сложности, потом он повышается.

Второй вопрос — это вопрос об уровнях сложности. Например, простую задачу на тему парной линейной регрессии можно усложнить следующими способами:

- использовать более объемный датафрейм, или датафрейм с шумовыми данными (добавляя этап очистки), или несколько датафреймов, которые необходимо соединить в один;

- использовать разные алгоритмы оценки параметров регрессии, разные схемы обучения (например, с кросс-валидацией);

- усложнять саму модель — включать возможность влияния факторов на коэффициенты линейной модели и т. д.

Очевидно, что в рамках одной темы и одного занятия охватить все возможные типы задач разной сложности нельзя, а это ставит вопрос о принципах построения блоков задач для нескольких занятий по одной теме.



Третий вопрос — это вопрос об экспериментах и анализе. Теоретическими рассуждениями возможно определить только основные принципы построения систем заданий по анализу данных. Вопрос же об эффективности разных систем заданий (и разных таксономий) может быть решен только экспериментальным путем.

Заключение

Представленное исследование, посвященное вопросам таксономий образовательных задач в области анализа данных для студентов высшего образования, позволило, с одной стороны, сформулировать авторский подход к построению конкретных задач и их систем, а с другой — обозначить большой круг вопросов, требующих дальнейшей проработки и исследований. Вне всяких сомнений, педагогика анализа данных в ближайшие годы станет актуальной темой исследований.

Список литературы

1. *Дорожная карта развития «сквозной» цифровой технологии «Нейротехнологии и искусственный интеллект»* / Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации. URL: <https://digital.gov.ru/ru/documents/6658/> (дата обращения: 15.06.2024).
2. *Тербушева Е. А.* Методика обучения студентов с профильной вузовской подготовкой в области математики интеллектуальному анализу данных : дис. ... канд. пед. наук. СПб., 2022.
3. *Салахова А. А.* Методика обучения основам искусственного интеллекта и анализа данных в курсе информатики на уровне среднего общего образования : дис. ... канд. пед. наук. М., 2022.
4. *Меренкова П. А.* Вариативное обучение системам искусственного интеллекта в рамках учебного предмета «Информатика» основной школы : дис. ... канд. пед. наук. М., 2023.
5. *Demchenko Y., Belloum A. S. Z., Los W., Wiktorski T. et al.* EDISON Data Science Framework: A Foundation for Building Data Science Profession For Research and Industry // 2016 IEEE international conference on cloud computing technology and science (CloudCom). Luxembourg, 2016. P. 620–626.
6. *Asamoah D. A., Doran D., Schiller S.* Interdisciplinarity in data science pedagogy: a foundational design // Journal of Computer Information Systems. 2020. Vol. 60, №4. P. 370–377.
7. *Ding Y., Park H., Lai Sh. et al.* Cross courses pedagogy development in data analytics // Proceedings of the 2021 ACM Southeast Conference. N. Y., 2021. P. 240–242.
8. *Aikat J., Carsey Th.M., Fecho K. et al.* Scientific training in the era of big data: a new pedagogy for graduate education // Big data. 2017. Vol. 5, №1. P. 12–18.
9. *Об утверждении профессионального стандарта «Специалист по большим данным»* : приказ Минтруда России от 06.07.2020 г. №405н (зарегистрирован в Минюсте России 05.08.2020 г. №59174). Доступ из справ.-правовой системы «КонсультантПлюс».
10. *Бортник Л. И., Кайгородов Е. В., Раенко Е. А.* О некоторых проблемах преподавания математики в высшей школе // Вестник Томского государственного педагогического университета. 2013. №4 (132). С. 19–24.



11. Пирогов В. Ю. Обучение программированию в высшей школе – проблемы и решения // Мир науки. Педагогика и психология. 2022. Т. 10, №6. С. 33.
12. Тестовое задание // github.com. URL: https://github.com/slgero/testovoe/blob/master/Revo:Mokka/data/revo_ds_test_task.csv (дата обращения: 23.06.2024).
13. Тестовое задание // github.com. URL: <https://github.com/kolya95/AccentureDSTestCase/blob/1c02aaf377d3c6ea73b4d5c8a2af97db11d20e9e/task2.txt> (дата обращения: 23.06.2024).
14. Бабикова Н.Н. Проектирование результатов обучения с использованием модифицированной таксономии Блума // Психология и педагогика: методика и проблемы практического применения. 2015. №46. С. 77–84.
15. Клушина Е.А. Активизация познавательной деятельности магистрантов в контексте таксономии Б. Блума // Вестник Северо-Кавказского федерального университета. 2016. №3. С. 156–160.
16. Бактыбаев Ж. Ш. Использование технологии таксономии Блума в учебном процессе вуза // Ярославский педагогический вестник. 2017. №1. С. 150–153.
17. Мусаева Н.Н., Авлиякулов Н.Х. Таксономия учебных целей для современного учебного процесса системы среднего специального и высшего профессионального образования // Проблемы современного образования. 2017. №3. С. 76–81.
18. Blyth W. A. L., Bloom B. S., Krathwohl D. R. Taxonomy of educational objectives. Handbook 1: Cognitive domain, taxonomy of educational objectives. Handbook 2: Affective domain. Rev. by W. A. L. Blyth // British Journal of Educational Studies. 1966. Vol. 14, №3. P. 119.
19. Masapanta-Carrión S., Velázquez-Iturbide J.Á. A systematic review of the use of bloom's taxonomy in computer science education // Proceedings of the 49th ACM technical symposium on Computer Science Education. N. Y., 2018. P. 441–446.
20. Whalley J., Lister R., Thompson E. et al. An Australasian study of reading and comprehension skills in novice programmers, using the Bloom and SOLO taxonomies // Proceedings of the 8th Australasian Conference on Computing Education. 2006. Vol. 52. P. 243–252.
21. Gluga R., Kay J., Lister R. et al. Over-confidence and confusion in using bloom for programming fundamentals assessment // Proceedings of the 43rd ACM technical symposium on Computer Science Education. N. Y., 2012. P. 147–152.
22. Anderson L.W., Krathwohl D.R. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition. Addison Wesley Longman, Inc., 2001.
23. Thompson E., Luxton-Reilly A., Whalley J.L. et al. Bloom's taxonomy for CS assessment // Proceedings of the 10th conference on Australasian Computing Education. 2008. Vol. 78. P. 155–161.
24. Meerbaum-Salant O., Armoni M., Ben-Ari M. Learning computer science concepts with scratch // Proceedings of the Sixth international workshop on Computing education research. 2010. P. 69–76.
25. Marzano R. J., Kendall J. S. (eds.). The new taxonomy of educational objectives. Corwin Press, 2006.
26. Фалина И.Н., Луговской К.И. Построение системы учебных задач с использованием таксономии Толлингеровой // Вестник Московского университета. Сер. 20: Педагогическое образование. 2015. №3. С. 97–107.
27. Clarkson P., Bishop A., Seah W.T. Mathematics education and student values: The cultivation of mathematical wellbeing // International research handbook on values education and student wellbeing. 2010. P. 111–135.



28. Irvine J. Taxonomies in education: Overview, comparison, and future directions // Journal of Education and Development. 2021. Vol. 5, №2. P. 1 – 26.

29. Переслегин С., Переслегина Е., Желтов А., Луковникова Н. Сумма стратегий. М., 2021.

Об авторе

Артем Дмитриевич Черемухин – канд. экон. наук, доц., Нижегородский государственный инженерно-экономический университет, Россия.

E-mail: ngieu.cheremuhin@yandex.ru

A. D. Cheremuhin

ON THE ISSUE OF CONSTRUCTING A TAXONOMY OF TASKS IN TEACHING DATA ANALYSIS

Nizhny Novgorod State University of Engineering and Economics,

Nizhny Novgorod, Russia

Received 18 June 2024

Accepted 27 October 2024

doi: 10.5922/vestnikpsy-2025-1-11

To cite this article: Cheremuhin A. D., 2025. On the issue of constructing a taxonomy of tasks in teaching data analysis, *Vestnik of Immanuel Kant Baltic Federal University. Series: Philology, Pedagogy, Psychology*, №1. P. 123 – 134. doi: 10.5922/vestnikpsy-2025-1-11.

The widespread use of artificial intelligence technologies, data analysis, and their positioning as the foundation for the development of the future economy significantly increases the demand for certain specialists. In this context, it is essential to focus on the methodology for teaching data analysis in universities. The aim of this article is to develop the basis of a taxonomy for creating practical data analysis tasks and to test its applicability. The author examines the main taxonomies of educational tasks, arguing that their limited applicability to the field of data analysis is due to its multidisciplinary and multifaceted nature. A horizontal-vertical taxonomy of educational tasks is proposed, based on three core processes: thinking, communication, and activity. The complexity of tasks is suggested as the key criterion for the vertical part of the taxonomy. Using the topic of "linear regression" as an example, typical tasks corresponding to primary and secondary processes, as well as their intersections, are presented. Specific assignments for Master's students in the "Business Informatics" program within the course "Computer Data Analysis" using the R language are developed and demonstrated. The need for further research in this area is justified, and a number of questions for future work are posed.

Keywords: taxonomy, learning task, data analysis, practical lesson

The author

Dr Artem D. Cheremuhin, Associate Professor, Nizhny Novgorod State University of Engineering and Economics, Russia.

E-mail: ngieu.cheremuhin@yandex.ru