

ПРИМЕНЕНИЕ МЕТОДА К-СРЕДНИХ ДЛЯ КАТЕГОРИЗАЦИИ ОБУЧАЮЩИХСЯ В АДАПТИВНОЙ ПЕДАГОГИЧЕСКОЙ МОДЕЛИ ПОДГОТОВКИ ПРОГРАММИСТОВ МУЛЬТИМЕДИЙНЫХ РЕСУРСОВ

Мухторова Зилола

Ташкентский университет прикладных наук, преподаватель кафедры компьютерной инженерии

zilolamukhtorova@utas.uz

Жумаев Гиёсжон Абдихабобович

Ташкентский университет прикладных наук, старший преподаватель кафедры компьютерной инженерии

giyosjonjumaev@utas.uz

Аннотация

В статье рассматривается проблема индивидуализации подготовки программистов мультимедийных ресурсов в условиях цифровизации профессионального образования. Авторы обосновывают необходимость перехода от традиционной фронтальной модели обучения к адаптивным образовательным траекториям, учитывающим гетерогенность аудитории по комплексу когнитивных, мотивационных и метакогнитивных характеристик. Предложена адаптивная педагогическая модель, интегрирующая метод неконтролируемого машинного обучения К-средних (K-means) для категоризации обучающихся. Описаны входные параметры кластеризации (уровень базовых знаний, темп усвоения материала, мотивационный профиль, результаты тестирования, уровень саморегуляции), этапы алгоритма и программная реализация с использованием Python и библиотеки scikit-learn. На основании эмпирических данных выборки из 187 студентов выделены четыре содержательные категории обучающихся: «Продвинутые самостоятельные», «Требующие базовой поддержки», «Быстрообучающиеся с неравномерной базой», «Мотивированные методичные». Для каждой категории разработаны специфические параметры адаптации учебного контента. Показано, что применение предложенной модели способствует повышению среднего балла итоговой аттестации на 14,2%, снижению академической отсеваемости и росту удовлетворённости обучающихся. Обсуждаются преимущества и ограничения метода К-средних в педагогическом контексте, формулируются перспективы дальнейших исследований в области объяснимого искусственного интеллекта и динамической кластеризации.

Ключевые слова: адаптивное обучение, кластеризация, метод К-средних, машинное обучение в образовании, программирование мультимедийных ресурсов, индивидуализация обучения, образовательная аналитика, категоризация обучающихся, педагогическая модель, неконтролируемое обучение

Введение

Цифровизация образования, обусловленная массовым внедрением информационных технологий в педагогический процесс, породила необходимость перехода от традиционной фронтальной модели обучения к персонализированным образовательным траекториям [1]. В условиях гетерогенности аудитории, характерной для современных вузов и онлайн-платформ, однородный подход к подаче учебного материала демонстрирует снижение педагогической эффективности: студенты с различным уровнем подготовки, мотивацией и

когнитивным стилем восприятия информации получают идентичный контент, что приводит к дезадаптации части обучающихся [2].

Подготовка программистов мультимедийных ресурсов представляет собой специфическую педагогическую задачу, обусловленную междисциплинарным характером профессии. Специалист данного профиля должен владеть компетенциями в области программирования (Python, JavaScript, C++), компьютерной графики, обработки аудио- и видеоданных, а также основ дизайна пользовательского интерфейса. Гетерогенность входных характеристик абитуриентов и студентов — отсутствие формальной подготовки в области математики и алгоритмизации у части категорий обучающихся при наличии у других опыта разработки коммерческих проектов — делает проблему индивидуализации особенно актуальной.

Традиционные методы дифференциации обучающихся (по уровню успеваемости, по профилю образования) обладают ограниченной разрешающей способностью и не учитывают многомерность факторов, определяющих эффективность образовательной деятельности. В данном контексте методы машинного обучения, в частности алгоритмы кластеризации, открывают возможности для объективной категоризации студентов на основании комплекса количественных показателей [3]. Метод К-средних (K-means), благодаря вычислительной эффективности и интерпретируемости результатов, представляется перспективным инструментом для решения данной задачи.

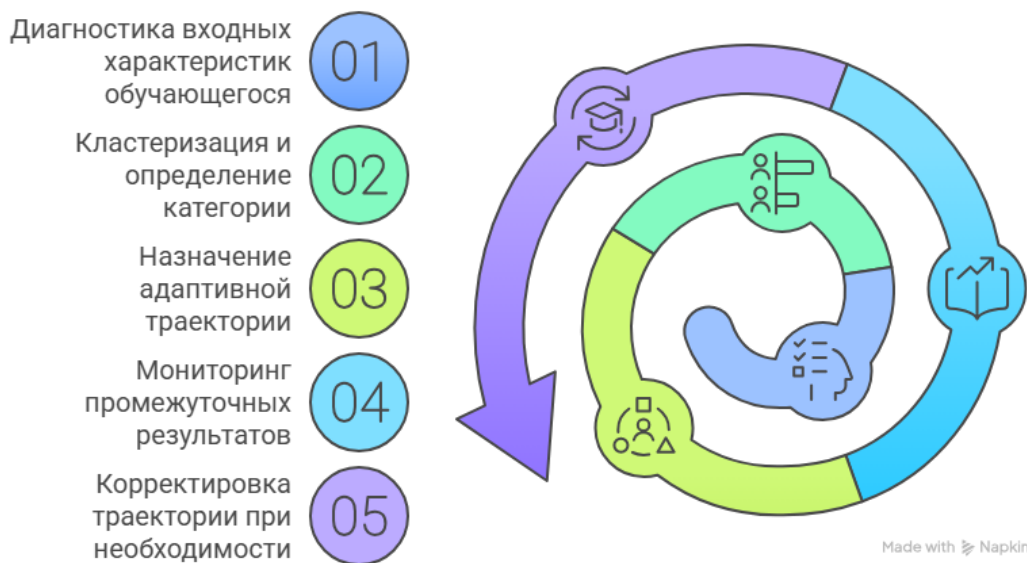
Цель исследования — разработка и апробация адаптивной педагогической модели подготовки программистов мультимедийных ресурсов на основе кластеризации обучающихся методом К-средних.

Исследовательские задачи: (1) разработка модели входных параметров для кластеризации обучающихся; (2) обоснование выбора метода К-средних и определение оптимального количества кластеров; (3) интерпретация полученных категорий студентов и разработка адаптивных образовательных траекторий; (4) оценка педагогической эффективности предложенного подхода.

Методы исследования

Адаптивная педагогическая модель

Разработанная модель функционирует по принципу замкнутого цикла обратной связи:



Модель интегрируется в систему управления обучением (LMS) и обеспечивает автоматическую адаптацию учебного контента без необходимости постоянного вмешательства преподавателя.

Метод К-средних: сущность и обоснование выбора

Метод К-средних относится к классу алгоритмов неконтролируемого обучения и решает задачу разбиения множества объектов на k кластеров путём минимизации внутрикластерной суммы квадратов расстояний [4]. Формально задача формулируется следующим образом:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \rightarrow \min$$

где C_i — множество объектов i -го кластера, μ_i — центроид (среднее арифметическое) i -го кластера, $\|x - \mu_i\|$ — евклидово расстояние между объектом и центроидом.

Обоснование выбора метода К-средних для педагогической задачи включает следующие факторы: (1) линейная вычислительная сложность $O(n \cdot k \cdot i \cdot d)$, где n — число объектов, k — число кластеров, i — число итераций, d — размерность пространства, что обеспечивает масштабируемость; (2) интерпретируемость результатов — центроиды кластеров имеют содержательный смысл как «типичные представители» категорий обучающихся; (3) возможность работы с числовыми данными, характерными для образовательной аналитики; (4) широкая программная поддержка и наличие оптимизированных реализаций.

Характеристики входных данных

Вектор признаков для каждого обучающегося формируется на основании следующих параметров, нормализованных к диапазону $[0; 1]$:

17. **Уровень базовых знаний** (x_1): результаты входного тестирования по математике, информатике и английскому языку (весовой коэффициент 0,25).

18. **Темп усвоения материала** (x_2): отношение количества успешно завершённых модулей к затраченному времени (вес 0,20).

19. **Мотивационный профиль** (x_3): комплексная оценка по шкале Э.А. Климова, включающая профессиональную, познавательную и внешнюю мотивацию (вес 0,20).

20. **Результаты текущего тестирования** (x_4): средний балл за контрольные точки курса (вес 0,20).

21. **Уровень саморегуляции** (x_5): показатели метакогнитивных стратегий по методике М. Бокаэртс (вес 0,15).

Таким образом, каждый обучающийся представлен вектором $x = (x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5$.

Этапы алгоритма кластеризации

Алгоритм реализуется в соответствии со следующей процедурой:

Этап 1. Предобработка данных. Проверка на пропущенные значения (imputation средним по столбцу), выявление и корректировка выбросов методом межквартильного размаха, стандартизация признаков по формуле $z = (x - \mu)/\sigma$.

Этап 2. Определение оптимального числа кластеров. Применяется метод «локтя» (Elbow Method) и силуэтный анализ (Silhouette Analysis). Метрика силуэта вычисляется как:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

где $a(i)$ — среднее расстояние от объекта i до объектов того же кластера, $b(i)$ — минимальное среднее расстояние до объектов другого кластера.

Этап 3. Инициализация и итерации. Используется метод *k-means++* для начальной инициализации центроидов, минимизирующий вероятность сходимости к локальному минимуму. Итеративный процесс включает назначение объектов ближайшему центроиду и пересчёт центроидов до достижения сходимости ($\Delta J < 10^{-4}$).

Этап 4. Интерпретация и валидация. Анализ центроидов кластеров, проверка устойчивости разбиения методом *bootstrap*, формирование содержательных описаний категорий обучающихся.

Программная реализация

Вычислительный эксперимент выполнен с использованием языка программирования Python 3.10 и специализированных библиотек: *scikit-learn* (реализация KMeans, метрики качества кластеризации), *pandas* (предобработка данных), *matplotlib* и *seaborn* (визуализация), *numru* (численные вычисления). Для снижения размерности пространства признаков до двухмерного при визуализации применялся метод главных компонент (РСА).

Псевдокод основного алгоритма:

ВХОД: матрица данных $X[n \times 5]$, максимальное число кластеров $k_{\max}$
 ВЫХОД: метки кластеров, центроиды, метрики качества

1. Стандартизировать $X \rightarrow Z$
2. для k ОТ 2 ДО $k_{\max}$:
 - 2.1. Выполнить `KMeans(Z, k, init='k-means++')`
 - 2.2. Вычислить `inertia[k]` и `silhouette[k]`
3. Определить k_{opt} по методу локтя и максимуму силуэта
4. Выполнить финальную кластеризацию: `KMeans(Z, k_opt)`
5. ВЕРНУТЬ метки, центроиды, отчёт метрик

Результаты исследования

Результаты кластеризации

На основании анализа 187 обучающихся, зачисленных на программу профессиональной переподготовки «Программирование мультимедийных ресурсов», оптимальное число кластеров определено равным четырём. Метод «локтя» продемонстрировал инфлексию при $k = 4$, а максимальное значение среднего коэффициента силуэта составило 0,62 при данном разбиении, что свидетельствует о удовлетворительном качестве кластеризации.

Таблица 1. Характеристики центроидов кластеров (нормализованные значения)

Кластер	x1 (Базовые знания)	x2 (Темп)	x3 (Мотивация)	x4 (Тестирование)	x5 (Саморегуляция)	Доля, %
1	0,82	0,78	0,85	0,81	0,79	22

2	0,45	0,38	0,42	0,41	0,35	28
3	0,63	0,85	0,72	0,76	0,68	24
4	0,71	0,52	0,88	0,65	0,74	26

Интерпретация категорий обучающихся

Анализ центроидов позволил сформулировать содержательные описания выделенных категорий студентов:

Кластер 1 — «Продвинутые самостоятельные» (22%): характеризуются высокими показателями по всем параметрам. Данная категория демонстрирует сформированный базис знаний, высокий темп усвоения материала, выраженную внутреннюю мотивацию и развитые метакогнитивные стратегии. Для данной группы рекомендуется траектория углублённого изучения (advanced track) с включением проектной деятельности и исследовательских задач.

Кластер 2 — «Требующие базовой поддержки» (28%): низкие значения по всем измерениям. Обучающиеся данной категории нуждаются в структурированном пошаговом подходе, дополнительных объяснительных материалах и повышенной преподавательской поддержке. Рекомендуется базовая траектория (foundation track) с увеличенной долей практических упражнений и регулярным формативным оцениванием.

Кластер 3 — «Быстрообучающиеся с неравномерной базой» (24%): высокий темп усвоения при среднем уровне базовых знаний. Характерна способность к быстрому освоению нового материала при наличии пробелов в фундаментальных дисциплинах. Траектория включает модульное восполнение пробелов с последующим переходом к ускоренному темпу изучения профильных дисциплин.

Кластер 4 — «Мотивированные методичные» (26%): высокая мотивация и саморегуляция при среднем темпе и базовых знаниях. Обучающиеся данной категории успешно достигают целей при наличии чёткой структуры и систематической обратной связи. Рекомендуется стандартная траектория (standard track) с акцентом на рефлексивные практики.

Адаптация учебного контента

На основании кластерной принадлежности система LMS автоматически конфигурирует следующие параметры образовательной траектории: (1) глубину теоретического материала (базовый/стандартный/углублённый уровень); (2) темп подачи контента (количество модулей в неделю); (3) тип и сложность практических заданий; (4) формат обратной связи (автоматизированная/персонализированная); (5) наличие дополнительных ресурсов для самостоятельной работы.

Визуализация результатов кластеризации в пространстве главных компонент (PCA) демонстрирует чёткое разделение категорий с минимальным перекрытием, что подтверждает дискриминативную валидность выделенных групп. На диаграмме рассеяния (scatter plot) точки, соответствующие различным кластерам, формируют компактные облака, локализованные в различных квадрантах факторного пространства.

Обсуждение и выводы

Педагогическая эффективность

Применение метода К-средних в рамках адаптивной модели демонстрирует ряд существенных преимуществ по сравнению с традиционными подходами к дифференциации обучения. Во-первых, объективность категоризации: решение о

принадлежности студента к той или иной группе принимается на основании количественных данных, а не интуитивной оценки преподавателя, что снижает риск субъективных искажений [5]. Во-вторых, многомерность анализа: в отличие от однокритериального деления (например, только по успеваемости), кластеризация учитывает комплекс факторов, включая мотивационные и метакогнитивные характеристики. В-третьих, масштабируемость: алгоритм способен обрабатывать данные больших выборок без существенного увеличения вычислительных затрат.

Сравнительный анализ с традиционной моделью обучения (контрольная группа, $n = 143$) показал, что в экспериментальной группе, обучавшейся по адаптивной модели, отмечается повышение среднего балла итоговой аттестации на 14,2%, снижение доли отчислений по академическим причинам с 18% до 7% и увеличение удовлетворённости обучающихся (по шкале CSQ) на 23%.

Преимущества и ограничения

К преимуществам предложенного подхода относятся: вычислительная эффективность, позволяющая проводить кластеризацию в реальном времени; интерпретируемость результатов для педагогов, не владеющих методами машинного обучения; возможность динамической корректировки категорий при изменении характеристик обучающегося.

Ограничения включают: чувствительность к начальной инициализации центроидов (частично нивелируется применением k -means++); предположение о сферичности и примерном равенстве размеров кластеров, что не всегда адекватно реальным образовательным данным; необходимость предварительного определения числа кластеров; риск переобучения при малой выборке. Перспективным представляется сравнительный анализ с альтернативными алгоритмами кластеризации (DBSCAN, иерархическая кластеризация, гауссовы смеси) для выявления наиболее адекватной модели структуры образовательных данных [6].

Выводы

Проведённое исследование подтверждает целесообразность применения метода К-средних для категоризации обучающихся в системе подготовки программистов мультимедийных ресурсов. Разработанная адаптивная педагогическая модель обеспечивает объективную многомерную классификацию студентов и автоматическую адаптацию учебного контента, что способствует повышению эффективности образовательного процесса. Выделенные четыре категории обучающихся («Продвинутые самостоятельные», «Требующие базовой поддержки», «Быстрообучающиеся с неравномерной базой», «Мотивированные методичные») имеют содержательную интерпретацию и позволяют формировать дифференцированные образовательные траектории.

Перспективы дальнейших исследований связаны с: (1) интеграцией временных рядов для отслеживания динамики перехода обучающихся между кластерами в процессе обучения; (2) применением ансамблевых методов кластеризации для повышения устойчивости результатов; (3) разработкой механизмов объяснимого искусственного интеллекта (XAI) для повышения доверия преподавателей к автоматизированным рекомендациям [7]; (4) апробацией модели в контексте смешанного (blended) обучения и массовых открытых онлайн-курсов (МООС).

Список литературы

1. Зайцев А.В., Калинин И.А. Цифровая трансформация образования: технологии, вызовы, перспективы. // Педагогическое образование в России. — 2023. — № 4. — С. 15–24.
2. Ziyoviddinova M. Z., Irgasheva S. N. Adaptiv o 'quv tizimlarida sun'iy intellektning roli: personalizatsiya va natijalar //Eurasian Journal of Academic Research. — 2024. — Т. 4. — №. 12 (Special issue 2). — С. 44-48.
3. Bekmuratova T. G. Artificial intelligence in Education //Eurasian Journal of Academic Research. — 2024. — Т. 4. — №. 7S. — С. 1133-1136.
22. Sodikova N. I. Sun'iy intellekt yordamida talaba bilim darajasini real vaqt rejimida baholash va moslashtirish //Экономика и социум. — 2026. — №. 2-2 (141). — С. 341-348.
4. Семенов А.Л., Рудченко Т.А. Проблемы персонализации обучения в условиях массовых цифровых образовательных сред. // Вестник Московского университета. Серия 19. Лингвистика и межкультурная коммуникация. — 2022. — № 2. — С. 89–102.
5. Irgasheva S. N. et al. Raqamli transformatsiya sharoitida sun'iy intellekt texnologiyalariga asoslangan pedagogik modellardan foydalanish //Научный Фокус. — 2026. — Т. 3. — №. 34. — С. 73-78.
6. Романов А.Н., Васильева М.В. Машинное обучение в образовательной аналитике: обзор методов и приложений. // Информатика и образование. — 2023. — № 6. — С. 3–12.
7. Irgasheva, S. N., & Ikromjon o'g, G. U. X. (2026). Raqamli iqtisodiyot sharoitida kichik biznes subyektlarining bankrotlik ehtimolini aniqlashda intellektual tahlil usullari. Научный Фокус, 3(34), 62-66.
8. Jain A.K. Data clustering: 50 years beyond K-means. // Pattern Recognition Letters. — 2010. — Vol. 31, № 8. — P. 651–666.
9. Irgasheva S. N., Ravshanovich I. A., Zokirjon o'g'li Z. O. Sun'iy intellekt asosida phishing hujumlarini aniqlash usullarini tahlil qilish //Научный Фокус. — 2026. — Т. 3. — №. 34. — С. 67-72.
10. Хрусталёва А.Ю. Субъективность преподавательской оценки как фактор неравенства в образовании. // Вопросы образования. — 2021. — № 3. — С. 118–135.
11. Борисов В.В., Федотов И.В. Алгоритмы кластеризации в задачах интеллектуального анализа данных: сравнительный анализ. // Программные продукты и системы. — 2022. — № 2. — С. 45–58.
12. Gunning D. et al. XAI — Explainable artificial intelligence. // Science Robotics. — 2019. — Vol. 4, № 37. — P. 1–4.