

УДК 004.942

doi 10.54708/22259309_2025_23365

КЛАСТЕРИЗАЦИЯ КАК ИНСТРУМЕНТ ПОИСКА ОБЪЕКТОВ-АНАЛОГОВ ПРИ РАЗРАБОТКЕ ПРОЕКТНОЙ ДОКУМЕНТАЦИИ НЕФТЯНЫХ МЕСТОРОЖДЕНИЙ

И. В. МАКАРОВ¹, В. Б. ПРУДНИКОВ²

¹makarov.uust@mail.ru, ²prudnikov.bgu@mail.ru

ФГБОУ ВО «Уфимский университет науки и технологий» (УУНИТ)

Аннотация. В статье рассматривается применение кластеризации по методу k-средних для поиска объектов-аналогов при разработке объектов нефтяных месторождений и проектной документации. Исследование основано на анализе данных 1490 объектов разработки, расположенных в различных регионах России, учитывая как статические, так и динамические параметры, включая количество пересмотров проектно-технической документации. Этот подход позволяет оценить влияние характеристик объектов на частоту пересмотра проектной документации и эффективность разработки. Выявленные закономерности могут служить основой для оптимизации процессов управления проектной документацией, что, в свою очередь, способствует повышению экономической эффективности разработки. Исследование направлено на углубление понимания взаимосвязей между характеристиками объектов и частотой пересмотров проектно-технической документации, что имеет важное значение при проектировании и управлении объектами разработки углеводородов.

Ключевые слова: кластеризация; метод k-средних; объекты-аналоги; разработка нефтяных месторождений; проектно-техническая документация; анализ данных.

ВВЕДЕНИЕ

Разработка проектно-технической документации (ПТД) для нефтяных месторождений – это сложный и затратный процесс, который требует высокой квалификации специалистов и значительных временных и финансовых ресурсов. В условиях растущей конкуренции и необходимости оптимизации затрат нефтяные компании сталкиваются с необходимостью внедрения современных методов, способствующих сокращению времени и ресурсов на разработку новых и пересмотр существующих ПТД [11]. Одним из таких методов является метод поиска объектов-аналогов. Этот подход позволяет существенно улучшить процесс проектирования, используя накопленные данные и опыт для принятия обоснованных решений. В результате снижается вероятность ошибок в проектировании, что повышает шансы на успешную реализацию новых проектов. Например, если новый объект имеет схожие характеристики с уже известным и успешно эксплуатируемым объектом, информация о последнем может быть использована при разработке нового ПТД.

Объект разработки представляет собой участок недр, на котором осуществляется добыча углеводородов. Он включает один или несколько продуктивных пластов, объединённых для разработки как единое целое. В рамках одного объекта разработки может разрабатываться несколько пластов, которые имеют схожие физико-химические свойства и расположены в одной зоне месторождения [12]. Под объектами-аналогами понимаются объекты, для которых уже ведется разработка и которые имеют схожие геологические, геофизические и эксплуатацион-

ные характеристики с новым или планируемым объектом. Применение метода аналогов позволяет не только повысить качество проектной документации и уменьшить риски, связанные с неопределенностью, но и способствует более быстрой реализации проектов, что является критически важным в условиях конкуренции.

Одним из традиционных способов поиска объектов-аналогов является метод экспертной оценки. Этот подход заключается в привлечении специалистов, обладающих необходимыми знаниями и опытом. Экспертная оценка позволяет оперативно получить обоснованные заключения от профессионалов, что значительно ускоряет процесс принятия решений. Тем не менее важно отметить, что, несмотря на высокую степень квалификации экспертов, данный метод не гарантирует абсолютной точности результатов [6].

Цель проведенного исследования заключалась в проведении кластеризации объектов разработки для поддержки планирования и проектирования новых объектов с учетом характеристик объектов-аналогов.

Для анализа была использована репрезентативная выборка данных, включающая 1490 объектов разработки дочерних обществ компании ПАО НК «Роснефть», расположенных в различных регионах Российской Федерации: Западная Сибирь, Восточная Сибирь, Центральная Россия и Южный федеральный округ [7].

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

В последнее время для решения задачи поиска аналогов активно применяются подходы, основанные на анализе данных и машинном обучении. В частности, применяются различные алгоритмы кластеризации данных [2]. Существует несколько методов кластеризации, которые могут быть применены для поиска объектов-аналогов нефтяных месторождений [2, 4]:

1. *Метод k-средних* (k-means) предполагает разделение данных на заранее определённое количество кластеров, при этом каждый объект принадлежит только к одному из них. Метод хорошо подходит для поиска объектов-аналогов, так как позволяет создавать группы со схожими геологическими и технологическими характеристиками.

2. *Иерархическая кластеризация* позволяет визуализировать отношения между различными объектами с помощью построения иерархической структуры данных от отдельных объектов до объединения их в более крупные кластеры.

3. *DBSCAN* определяет принадлежность объектов кластерам, опираясь на характеристику «плотности точек». При этом используются два параметра: радиус и минимальное количество точек в кластере. Объекты, которые находятся в пределах одного радиуса, считаются принадлежащими одному кластеру. Этот метод позволяет выявлять скрытые структуры в данных.

Одним из наиболее популярных и эффективных методов кластеризации является классический метод k-средних, который широко применяется в различных областях, включая анализ данных, машинное обучение и научные исследования [13]. Выбор обусловлен его высокой эффективностью в разделении данных на заранее определённое количество кластеров, что позволяет группировать объекты с аналогичными геологическими и технологическими характеристиками. Кроме того, алгоритм k-средних легко адаптируется к различным наборам данных и позволяет изменять количество кластеров в зависимости от специфики анализа. По этой причине в настоящем исследовании для кластеризации нами был использован алгоритм k-средних.

Для эффективной кластеризации объектов разработки необходимо учитывать как статические, так и динамические параметры [3, 6]. Статические параметры, такие как общая толщина пласта и коэффициент пористости, описывают физические характеристики объекта и остаются неизменными во времени, в то время как динамические параметры отражают изменения, происходящие в процессе, и позволяют оценить эффективность эксплуатации месторождений.

В отличие от предыдущих исследований [1, 2, 4] данная работа рассматривает кластеризацию как один из способов оптимизации разработки проектно-технической документации (ПТД). При этом в качестве признаков кластеризации, помимо традиционных основных характеристик объектов, рассматривается количество произошедших пересмотров проектно-

технической документации (ПТД) по объекту, что позволяет оценить динамику разработки объекта в целом и повысить точность проектирования. Проектная документация представляет собой ключевой инструмент регулирования процесса разработки месторождений, обеспечивая не только методологическую основу для их освоения, но и выполняет функцию государственного контроля. В рамках ПТД устанавливаются допустимые отклонения для ключевых показателей, таких как объемы добычи нефти и газа, действующего фонда добывающих и нагнетательных скважин, ввода новых скважин, бездействующего фонда скважин, а также ряда других параметров [10]. Превышение этих отклонений может свидетельствовать о нестабильности проектных решений, принятых на этапе планирования, либо может указывать на изменение условий эксплуатации. В таких случаях возникает необходимость пересмотра проектной документации для приведения проекта в соответствие с требованиями контролирующих органов и сохранения действующих лицензий. Такой подход отличается от ранее проведенных исследований, в которых частота пересмотров ПТД не рассматривается как показатель эффективности разработки проектных документов. В результате становится возможной оценка влияния различных характеристик объектов на необходимость корректировок ПТД.

Таблица 1

Основные параметры объектов и их характеристики

Параметр	Описание	Диапазон	Среднее значение	
Статические	Общая толщина пласта, м	Расстояние от подошвы до кровли пласта	1 – 328	39
	Эффективная нефтенасыщенная толщина пласта, м	Суммарная толщина пласта, в которой содержится нефть. Отражает реальный объем нефти, доступный для извлечения	0,5 – 99	8,5
	Коэффициент пористости, доли ед.	Отражает способность пласта удерживать нефть. Высокий коэффициент пористости говорит о высокой емкости пласта для хранения углеводородов, что делает его более привлекательным для разработки	0,001 – 2,63	0,32
	Коэффициент песчанистости, доли ед.	Характеризует содержание песчаных частиц в пласте. Высокий коэффициент песчанистости указывает на лучшие условия для перемещения углеводородов по пласту, что способствует более эффективной добыче	0,01 – 1,12	0,55
	Коэффициент расчленённости, доли ед.	Определяет степень фрагментации пласта. Высокий уровень расчленённости может создать дополнительные барьеры для извлечения нефти, так как фрагментация породы усложняет движение жидкости	0,047 – 87	7,85
	Начальная пластовая температура, °С	Температура нефти в пласте до начала разработки месторождения. Более высокая температура способствует снижению вязкости нефти, облегчая её перемещение по пласту, что положительно сказывается на добыче	9 – 656	109
	Начальное пластовое давление, МПа	Давление нефти в пласте до начала разработки месторождения. От него зависит способность нефти выделяться естественным образом	4 – 219	44,6
	Содержание серы в нефти, %	Процентное содержание серы в общем объеме нефти. Высокое содержание серы требует дополнительных затрат на переработку	0,01 – 15,4	2,31
	Начальные извлекаемые запасы нефти (НИЗ), тыс. тонн	Объем нефти, который может быть извлечен при существующих технологиях и экономических условиях	1 – 97 108	3 075
Динамические	Текущие извлекаемые запасы нефти, тыс. тонн	Объем нефти, который остается в пласте после начала разработки	1 – 87 751	1 939
	Накопленная добыча нефти, тыс. тонн	Объем нефти, который был извлечен из пласта на данный момент	0 – 96 000	1 069
	Количество пересмотров ПТД	Число раз, когда проектная документация была создана или пересмотрена для объекта разработки	1 – 7	2

Источник: Составлено авторами на основе набора данных.

В рамках исследования был сформирован набор данных по 1490 объектам разработки, содержащий 9 статических и 3 динамических параметра. Отбор объектов проводился на основе тщательного анализа действующих проектно-технических документов (ПТД) на 2024 год. В выборку были включены объекты из различных регионов, таких как Западная Сибирь, Восточная Сибирь, Центральная Россия и Южный федеральный округ. Включенные в выборку объекты отличаются различным геологическим строением, типами залежей и характеристиками пластов, включая расчлененность, проницаемость, глубину залегания и объем запасов. В выборке были учтены как крупные, так и средние, и мелкие объекты, что позволяет учитывать различия в сложности разработки, зависящие от масштабов и объемов углеводородных запасов. Примененный стратифицированный метод отбора позволяет обеспечить репрезентативность выборки, учитывая разнообразие геологических и технологических характеристик объектов разработки. Данные параметры использованы нами в качестве признаков кластеризации, поскольку они играют ключевую роль в выборе технологий разработки, что, в свою очередь, влияет на эффективность и рентабельность проектов [6]. В табл. 1 приведены перечень выбранных параметров, а также описательная статистика по выборке.

РЕЗУЛЬТАТЫ

Перед кластеризацией объектов был проведен эксплораторный анализ данных. В первую очередь данные были проверены на наличие выбросов и пропусков. Пропущенных значений в исходных данных обнаружено не было. Для выявления аномалий и особенностей, таких как выбросы или ошибки в данных, были построены графики оценок функций плотности распределений. Эти графики позволили оценить основные статистические параметры, такие как среднее значение, медиана, стандартное отклонение, наличие выбросов, также визуально установить соответствие значений нормальному закону распределения. Графики оценок функции плотности распределения для каждой переменной представлены на рис. 1.

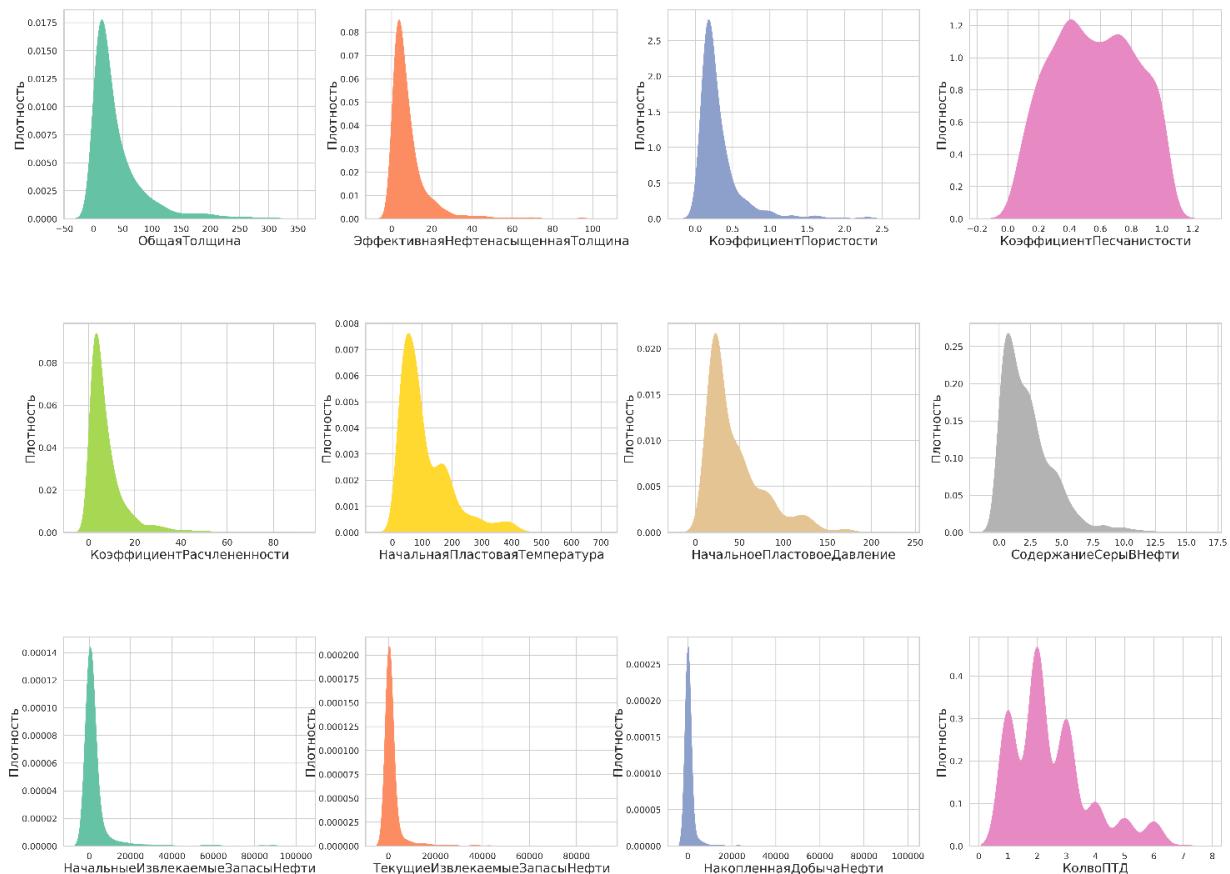


Рис. 1. Графики оценок функции плотности распределения переменных

В результате исследования было установлено, что параметры объектов разработки находятся в широком диапазоне и имеют различный масштаб. Распределение большей части переменных (общая толщина, пористость, начальное давление, извлекаемые запасы и накопленная добыча) является смещенным влево. Показатель песчанистости демонстрирует более равномерное распределение, что может указывать на вариативный состав пород. Начальная пластовая температура имеет два выраженных пика, что может свидетельствовать о геологических зонах с разными температурными условиями. Большинство объектов характеризуются низким содержанием серы, что положительно влияет на её качество и снижает коррозионные риски. Для приведения данных к единому масштабу был применён метод стандартизации *fit_transform*, доступный в библиотеке *sklearn*.

Оценки значений корреляционной матрицы и соответствующие р-значения указаны в скобках и представлены в табл. 2. Большая часть оценок являются статистически значимыми (р-значения $< 0,05$). В целом коэффициенты корреляции между признаками находятся в диапазоне от -0,1 до 0,7, что позволяет сделать вывод об отсутствии мультиколлинеарности. Таким образом, для кластеризации были отобраны все анализируемые переменные. Мы не использовали метод главных компонент (PCA) для уменьшения размерности признакового пространства, поскольку количество признаков относительно мало (12), что позволяет повысить интерпретируемость результатов в части характеристик объектов каждого из сформированных кластеров.

Таблица 2

Матрица корреляций и р-значений

	Общая толщина	Эффективная нефтенасыщенная толщина	Коэффициент пористости	Коэффициент песчанистости	Коэффициент расчлененности	Начальная пластовая температура	Начальное пластовое давление	Содержание серы в нефти	Начальные извлекаемые запасы нефти	Текущие извлекаемые запасы нефти	Накопленная добыча нефти	Количество пересмотротов ПГД
Общая толщина	1											
Эффективная нефтенасыщенная толщина	0,652 (0,000)	1										
Коэффициент пористости	0,306 (0,000)	0,255 (0,000)	1									
Коэффициент песчанистости	-0,174 (0,000)	0,110 (0,000)	0,176 (0,000)	1								
Коэффициент расчлененности	0,685 (0,000)	0,481 (0,000)	0,358 (0,000)	-0,092 (0,000)	1							

Продолжение таблицы 2

	Общая толщина	Эффективная нефтенасыщенная толщина	Коэффициент пористости	Коэффициент песчанистости	Коэффициент расщленности	Начальная пластовая температура	Начальное пластовое давление	Содержание серы в нефти	Начальные извлекаемые запасы нефти	Текущие извлекаемые запасы нефти	Накопленная добыча нефти	Количество пересмотров ПГП
Начальная пластовая температура	0,677 (0,000)	0,493 (0,000)	0,387 (0,000)	-0,103 (0,000)	0,925 (0,000)	1						
Начальное пластовое давление	0,626 (0,000)	0,447 (0,000)	0,369 (0,000)	-0,110 (0,000)	0,822 (0,000)	0,909 (0,000)	1					
Содержание серы в нефти	0,037 (0,158)	0,018 (0,484)	0,009 (0,717)	-0,018 (0,484)	0,009 (0,735)	0,036 (0,165)	0,125 (0,000)	1				
Начальные извлекаемые запасы нефти	0,069 (0,008)	0,042 (0,103)	0,023 (0,374)	0,004 (0,877)	0,074 (0,004)	0,091 (0,000)	0,024 (0,350)	-0,071 (0,006)	1			
Текущие извлекаемые запасы нефти	0,042 (0,106)	0,018 (0,477)	0,030 (0,241)	0,021 (0,417)	0,066 (0,011)	0,075 (0,004)	0,030 (0,248)	-0,108 (0,000)	0,810 (0,000)	1		
Накопленная добыча нефти	0,057 (0,029)	0,031 (0,238)	0,006 (0,809)	-0,021 (0,413)	0,045 (0,080)	0,063 (0,016)	0,005 (0,835)	0,005 (0,836)	0,698 (0,000)	0,152 (0,000)	1	
Количество ПГД	0,319 (0,000)	0,276 (0,000)	0,544 (0,000)	0,105 (0,000)	0,359 (0,000)	0,391 (0,000)	0,362 (0,000)	-0,008 (0,766)	0,040 (0,119)	0,034 (0,187)	0,027 (0,299)	1

Для корректной работы алгоритма k-средних и получения интерпретируемых результатов кластеризации необходимо предварительно определить оптимальное количество кластеров. Традиционно для этого используются методы построения дендрограммы с помощью иерархической кластеризации и «метод локтя» [20]. При построении иерархической дендрограммы был использован метод минимальной дисперсии Уорда [14]. Согласно методу, на каждом шаге происходит слияние пары кластеров, определяемых на основе оптимального значения целевой функции как суммы квадратов ошибок после их объединения. Тем самым метод Уорда способствует формированию кластеров с более однородной дисперсией характеристик объектов внутри кластера. В результате оптимальное количество кластеров было установлено равным трем (рис. 2).

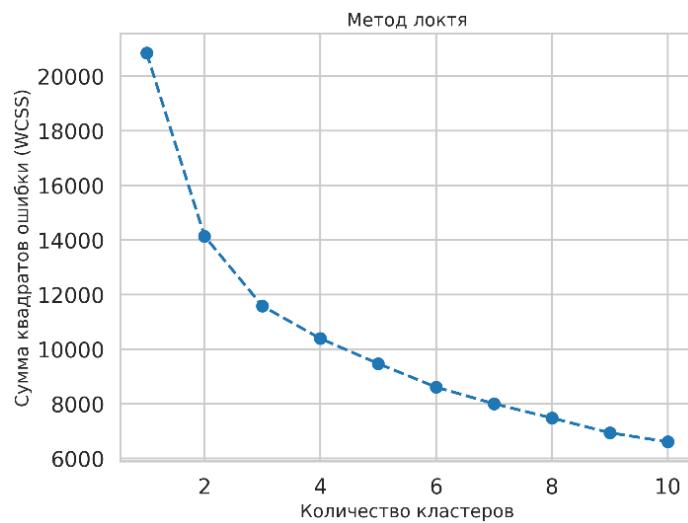


Рис. 2. Определение оптимального количества кластеров методом локтя

Для кластеризации мы использовали вариант алгоритма k-средних, предложенный Ч. Эльканом (C. Elkan) [15] и реализованный в классе `sklearn.cluster.KMeans`, со следующими параметрами: количество кластеров (`n_clusters`) – 3, число запусков (начальных положений центроидов, `n_init`) – 10, максимальное число итераций (`max_iter`) – 300, погрешность (`tol`) – 0.0001 [14].

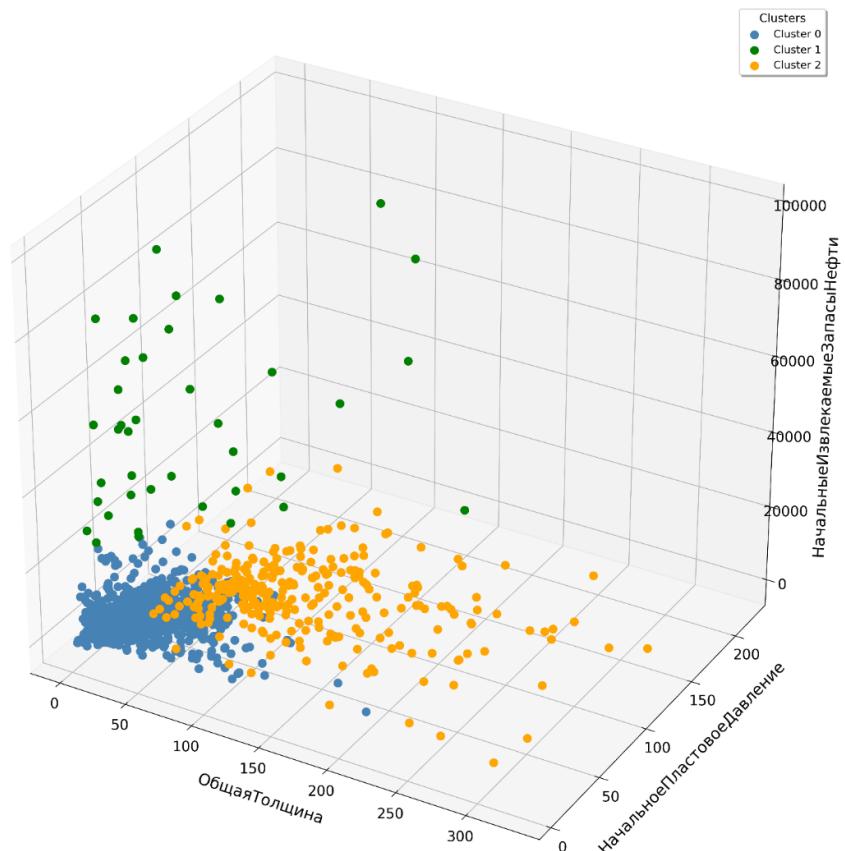


Рис. 3. 3D-визуализация кластеризации объектов разработки в координатных осях «Общая толщина», «Начальное пластовое давление» и «Начальные извлекаемые запасы нефти»

В результате кластеризации были выделены три кластера, которые в значительной степени различаются по средним значениям параметров входящих в них объектов (табл. 3). Кластер 0 объединяет объекты с минимальными значениями по всем трем параметрам. Кластер 1 включает объекты с высокими начальными извлекаемыми запасами нефти. Кластер 2 отличается равномерным распределением значений по всему графику. 3D-визуализация кластеров приведена на рис. 3 (в координатных осях «Общая толщина», «Начальное пластовое давление» и «Начальные извлекаемые запасы нефти») и рис. 4 (в координатных осях «Количество пересмотров ПТД», «Начальная пластовая температура» и «Коэффициент песчанистости»).

Объекты кластера 0 образуют более плотную группу точек, что указывает на стабильные геологические и технологические условия. Среднее число ПТД по объектам данного кластера минимально (табл. 3). Кластер 1 представлен более редкими объектами, что подчеркивает их уникальные характеристики и требует индивидуального подхода к разработке. Кластер 2 включает объекты с максимальным средним числом пересмотров ПТД, что может указывать на сложные геологические условия, требующие частых корректировок проектной документации.

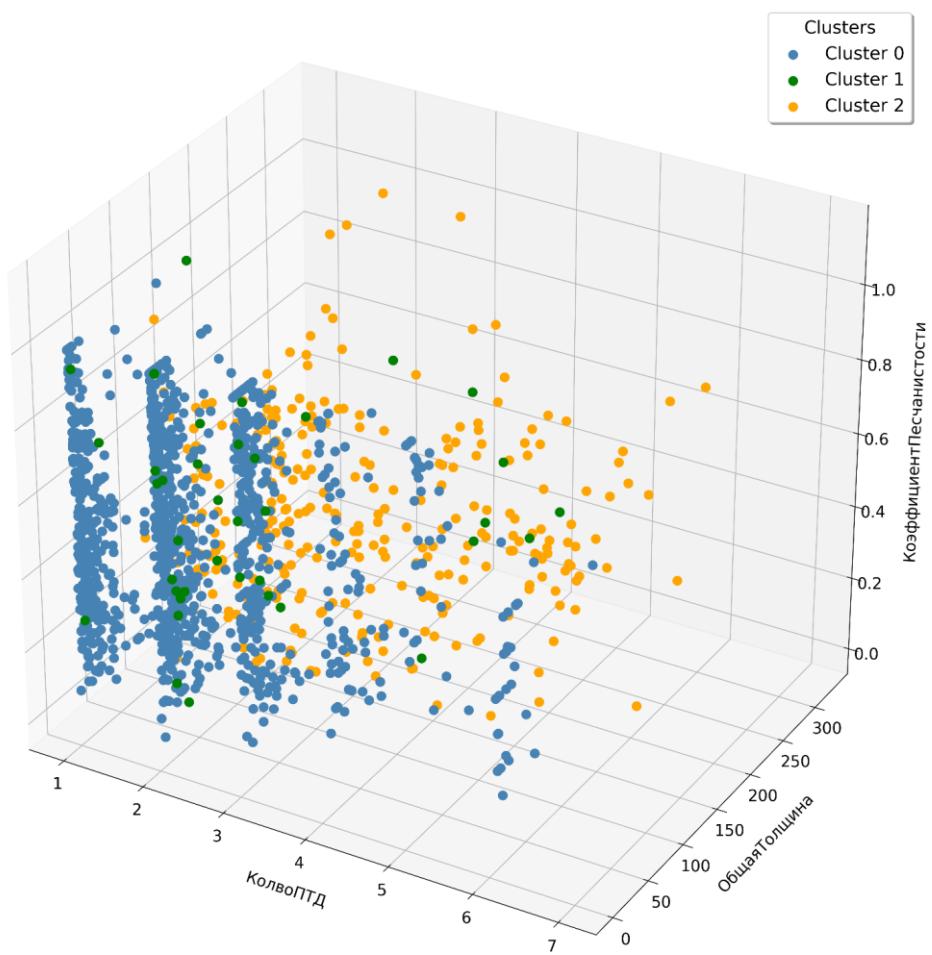


Рис. 4. 3D-визуализация кластеризации объектов разработки в координатных осях: количество пересмотров проектно-технической документации (ПТД), начальная пластовая температура и коэффициент песчанистости

Для описания кластеров мы опирались на средние значения признаков объектов, входящих в каждый из кластеров (табл. 3), а также на построенные 3D-визуализации кластеризации объектов разработки (рис. 3–4) и графики оценок плотности распределения (KDEPlot, Kernel Density Estimation Plot (рис. 5)). Характеристика кластеров приводится с позиции продуктивности и риска проведения разработки. Под риском в данном случае мы понимаем количество пересмотров

ПТД, которое характеризует расхождение (как в лучшую, так и в худшую сторону) между плановыми и фактическими показателями эксплуатации объектов разработки.

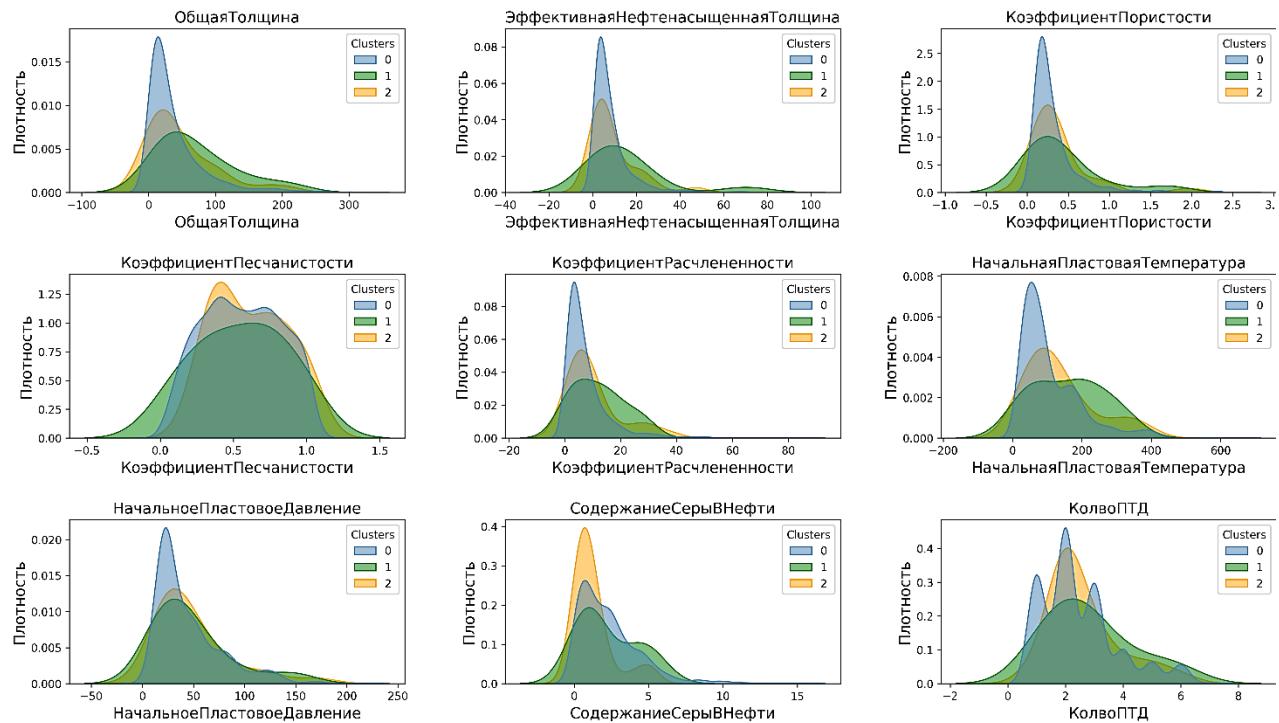


Рис. 5. Распределение кластеров по плотности

Таблица 3

Типовой портрет кластеров

Параметр	Кластер 0	Кластер 1	Кластер 2
Количество объектов в каждом кластере	1176	30	284
Общая толщина пласта, метров	23,66	102,72	57,05
Эффективная нефтенасыщенная толщина пласта, метров	5,55	19,94	11,98
Коэффициент пористости, доли ед.	0,26	0,57	0,42
Коэффициент песчанистости, доли ед.	0,56	0,55	0,59
Коэффициент расчлененности, доли ед.	4,7	20,16	10,41
Начальная пластовая температура, °С	75,78	240,26	140,49
Начальное пластовое давление, мПа	32,26	93,36	48,29
Содержание серы в нефти, %	2,26	2,63	1,61
Начальные извлекаемые запасы нефти, тыс. тонн	1803,02	2166,70	61410,17
Текущие извлекаемые запасы нефти, тыс. тонн	1205,10	1482,76	34932,70
Накопленная добыча нефти, тыс. тонн	559,61	643,53	25028,73
Количество ПТД, шт.	2,19	3,54	2,87

Первый, наиболее многочисленный, кластер 0 характеризуется наименьшими значениями по всем ключевым параметрам среди рассматриваемых кластеров. Низкая общая и эффективная нефтенасыщенная толщина пласта ограничивает его потенциал для извлечения углеводородов. Несмотря на относительную простоту разработки и умеренные значения температуры и давления, низкие коэффициенты пористости (0,26) и начальные запасы нефти делают этот кластер менее привлекательным для разработки. Тем не менее меньшее среднее число пересмотров ПТД (2,19) свидетельствует о том, что разработка этого класса может быть проще и более экономически выгодной. Объекты данного кластера следует охарактеризовать как «объекты с невысоким потенциалом добычи и относительно низким риском».

Кластер 1 выделяется наибольшей общей и эффективной толщиной пласта, что делает его наиболее перспективным для разработки. Высокий коэффициент пористости (0,57) в этом кластере способствует лучшему удержанию и извлечению нефти. Однако высокая начальная пластовая температура и давление могут указывать на более сложные условия разработки. Большое число пересмотров ПТД (3,54) указывает на сложность разработки и необходимость тщательного планирования и реализации проектов по объектам в данном кластере. Отметим, что число объектов, отнесенных ко второму кластеру, составляет менее 3 % от всей выборки. Объекты данного кластера следует охарактеризовать как «высокорисковые и потенциально максимально перспективные».

Третий кластер 2 включает объекты с умеренными значениями начальной пластовой температуры и давления, а также с высокими начальными запасами нефти, которые значительно превышают запасы в кластерах 0 и 1. Это делает данный класс потенциально выгодным для разработки. Однако низкие коэффициенты пористости и песчанистости могут потребовать применения дополнительных технологий для оптимизации процесса добычи. Умеренное число пересмотров ПТД (2,87) свидетельствует о средней сложности разработки данного класса. Объекты данного кластера следует охарактеризовать как «умеренно рисковые и продуктивные».

ЗАКЛЮЧЕНИЕ

Для кластеризации с целью определения объектов-аналогов были проанализированы 1490 объектов разработки нефтяных месторождений, принадлежащих обществам группе компаний ПАО НК «Роснефть», расположенных в различных регионах России: Западной и Восточной Сибири, Центральной России и Южном федеральном округе. Анализ объектов разработки проводился на основе действующих проектно-технических документов, разработанных в период с 2019 по 2024 год. Репрезентативность выборки была обеспечена методами стратифицированного отбора на этапе подготовки исходных данных. В качестве параметров объектов разработки рассмотрены, в частности, общая толщина пласта, эффективная нефтенасыщенная толщина пласта, коэффициенты пористости, песчанистости и расчленённости, начальное пластовое давление и температура, запасы добычи и др.

В результате применения алгоритма k-средних (вариация Ч. Элькана) были выделены три основных кластера объектов разработки с различными характеристиками:

- Кластер 0 – объекты с «невысоким потенциалом добычи и относительно низким риском» составляют наиболее многочисленную группу и характеризуются предсказуемыми условиями разработки;
- Кластер 1 – «высокорисковые и потенциально максимально перспективные» объекты, встречающиеся реже, но обладающие значительным потенциалом и требуют более индивидуального подхода к проектированию;
- Кластер 2 – «умеренно рисковые и продуктивные» объекты, которые характеризуются стабильными условиями разработки и требуют минимального количества пересмотров проектной документации.

Особенностью исследования является включение в анализ количества пересмотров проектно-технической документации по объектам разработки. Как показывает анализ, количество пересмотров ПТД существенно зависит от характеристик объектов, определяемых разнообразием географических и климатических условий. Объекты с высокими запасами и сложной структурой требуют большего числа пересмотров, что указывает на необходимость более тщательного подхода к их проектированию. Напротив, объекты с меньшими запасами и простой структурой требуют менее частых пересмотров, что может быть связано с более предсказуемыми условиями разработки.

Репрезентативность выборки позволяет распространить полученные выводы к объектам разработки, ПТД по которым еще только формируется, для поиска объектов-аналогов и по-

тенциальных рисков эксплуатации. Однако для дополнительной верификации результатов целесообразно расширить как объем выборки, так и набор параметров, которые потенциально могут быть связаны с частотой пересмотров проектно-технической документации.

СПИСОК ЛИТЕРАТУРЫ

1. **Безбородов А. К., Деева И. Ю.** Алгоритм поиска месторождений-аналогов на основе методов машинного обучения // Альманах научных работ молодых ученых Университета ИТМО. СПб: НИУ ИТМО, 2021. С. 1–6.
2. **Безбородов А. К., Деева И. Ю.** Поиск месторождений-аналогов на основе кластеризации байесовских сетей // Изв. вузов. Приборостроение. 2022. Т. 65. № 1. С. 64–72. DOI: 10.17586/0021-3454-2022-65-1-64-72.
3. **Вологин И. С., Исламов Р. Р., Нигматуллин Ф. Н., Харисова А. В., Лознюк О. А.** Методика выбора объекта-аналога для нефтегазовой залежи по геолого-физическими характеристикам // Нефтяное хозяйство. 2019. № 12. С. 124–127. DOI: 10.24887/0028-2448-2019-12-124-127. EDN: NRPKVE.
4. **Исламов Р. Р., Абдрахманова Э. К., Ялаев А. В., Нигматуллин Ф. Н., Муслимов Б. Ш., Харисова А. В.** Новая методика аналитического экспресс-расчета основных показателей разработки неразрабатываемых нефтегазовых залежей для проведения многовариантных расчётов с целью комплексной оптимизации проектных решений // Проблемы сбора, подготовки и транспорта нефти и нефтепродуктов. 2023. Вып. 3 (143). С. 48–60. <http://doi.10.17122/ntj-oil-2023-3-48-60>.
5. **Мухаметшин В. Ш., Хакимзянов И. Н., Бахтизин Р. Н., Кулешова Л. С.** Дифференциация и группирование сложно-построенных залежей нефти в карбонатных коллекторах в решении задач управления разработкой // Нефтяное хозяйство. 2021. № 5. С. 58–64.
6. **Орлова И. О., Захарченко Е. И., Скиба Н. К.** Методический подход к классификации месторождений и поиску месторождений-аналогов // Геология, геофизика и разработка нефтяных и газовых месторождений. 2014. № 12. С. 16–18.
7. Официальный сайт ПАО НК «Роснефть». Операционная структура. URL: <https://udmurtneft.rosneft.ru/about/Glance/OperationalStructure> (дата обращения: 09.10.2024).
8. **Перминов Д. Е., Валеев С. В.** Кластерный анализ с использованием элементов нечеткой логики с целью автоматического поиска скважин-кандидатов для проведения геолого-технических мероприятий // Научно-технический вестник ОАО «НК «Роснефть». 2019. № 12. С. 1–7.
9. **Поднебесных А. В., Хафизов А. Р.** Методика экспресс-оценки выбора объекта-аналога для залежей углеводородного сырья на основе их геологических признаков // Проблемы сбора, подготовки и транспорта нефти и нефтепродуктов. 2019. № 3. С. 9–18.
10. Правила подготовки технических проектов разработки месторождений углеводородного сырья, утвержденные приказом Минприроды России от 20.09.2019 № 639, в ред. приказов Минприроды России от 07.08.2020 № 570, от 06.10.2020 № 772 // Официальный сайт Министерства природных ресурсов и экологии Российской Федерации. URL: <http://www.mnr.gov.ru> (дата обращения: 20.10.2024).
11. **Ронжин А. А.** Проектная документация при разработке нефтегазоконденсатных месторождений // Молодой ученый. 2020. № 17 (307). С. 413–415.
12. **Росляк А. Т., Санду С. Ф.** Разработка и эксплуатация нефтяных и газовых месторождений: рекомендовано в качестве учебного пособия редакционно-издательским советом Томского политехнического университета // Томский политехнический университет. Томск: Изд-во Томского политехнического университета, 2013.
13. **Солодов И. С., Шакшин В. П., Колесников В. А., Хамитов И. Г., Устинов А. С.** Статистические подходы к выявлению месторождений-аналогов Самарской области // Нефтяное хозяйство. 2011. № 6. С. 30–33.
14. About K-Means vs Hierarchical Clustering Methods for Data Segmentation. Available at: <https://dataheadhunters.com/academy/k-means-vs-hierarchical-clustering-methods-for-data-segmentation/> (дата обращения: 19.09.2024).
15. **Elkan C.** Using the triangle inequality to accelerate k-means // ICML 2003. Conference Proceedings. P. 147–153.
16. Clustering. Available at: <https://r.qcbs.ca/workshop09/book-en/clustering.html> (дата обращения: 13.10.2024).
17. Da Gama e Silva R. et al. Sensitivity analysis in a machine learning methodology for reservoir analogues // Rio Oil & Gas Expo and Conf., Rio de Janeiro, Brazil, 24 – 27 Sept. 2018. P. 1–9.
18. **Neves L., Ferreira C., Vesely F.** Facies, petrophysical and geochemical properties of gravity-flow deposits in reservoir analogs from the Itararé Group (late Carboniferous), Paraná Basin, Brazil // Marine and Petroleum Geology. 2019. Vol. 110. P. 717–736. DOI: 10.1016/j.marpetgeo.2019.07.038.
19. **Olivo M. et al.** Characterizing scales of sedimentary heterogeneity in a fluvial–estuarine reservoir analog: An example from the Mulichinco Formation, Neuquén Basin, Argentina // Marine and Petroleum Geology. 2020. Vol. 121. DOI: 10.1016/j.marpetgeo.2020.104604
20. **Popova O.** Analogy in the World of Geological Uncertainties, or How Reservoir Analogs May Refine Your Probabilistic Geomodel // SPE Annual Caspian Techn. Conf. and Exhibition, Astana, Kazakhstan, 31 Oct. – 1 Nov. 2018. P. 1–13. DOI: 10.2118/192551-MS.
21. **Ren Z., Sun L., Zhai Q.** Improved k-means and spectral matching for hyperspectral mineral mapping // Intern. Journal of Applied Earth Observation and Geoinformation. 2020. Vol. 91. P. 1–12. DOI: 10.1016/j.jag.2020.102154.

22. Rodriguez H. M. et al. New Approach to Identify Analogue Reservoirs // SPE Economics and Management, New Orleans, Louisiana, USA, 30 Sept. – 2 Oct. 2013. P. 1–17. DOI: 10.2118/166449-MS.
23. Sun S. Q., Wan J. C. Geological analogs usage rates high in global survey // Oil & Gas Journal. 2002. Vol. 100, issue 46. P. 49–50.
24. Voskresenskiy A. et al. Variations in ranked list of reservoir analogs as an effect of search preferences // Geosciences: Converting Knowledge into Resources, Saint-Petersburg, Russia, 6 – 9 Apr. 2018. P. 1–5. (Conf. Proc.) DOI: 10.3997/2214-4609.202053163.

ОБ АВТОРАХ

МАКАРОВ Иван Владимирович, аспирант каф. СиБИ
ПРУДНИКОВ Вадим Борисович, д/н, доцент СиБИ.

METADATA

Title: Clustering as a tool for searching for analog objects In the development of oil field design documentation

Author: I.V. Makarov¹, V.B. Prudnikov²

Affiliation:

^{1,2} Ufa University of Science and Technology (UUST), Russia.

Email: ¹makarov.uust@mail.ru, ²prudnikov.bgu@mail.ru

Language: Russian.

Source: Molodezhnyj Vestnik UGATU (scientific journal of Ufa University of Science and Technology), no. 2 (33), pp. 65–76, 2025.

ISSN 2225-9309 (Print).

Abstract: The article applies clustering by the k-means method to search for similar objects in the development of oil fields and design documentation. The study is based on the analysis of data from 1490 objects from various regions of Russia, taking into account both static and dynamic parameters, including the number of revisions of design and technical documentation. This approach allows evaluating the characteristics of objects at the stages of revising design documentation and developing efficiency. The identified rules can serve to optimize the management processes of design documentation, which, in turn, contributes to increasing the economic efficiency of development. Research is aimed to deepen the understanding of the relationship between objects and the termination of revision of design and technical documentation, which is of critical importance in the design and management of hydrocarbon development facilities.

Key words: clustering, k-means method, analog objects, oil field development, design and technical documentation, data analysis.

About authors:

МАКАРОВ Иван Владимирович, postgraduate student, Dept. of STATISTICS AND BUSINESS INFORMEATICS (UUST).

ПРУДНИКОВ Вадим Борисович, docent, Dept. of STATISTICS AND BUSINESS INFORMEATICS (UUST).