

## **Аннотация**

В статье представлен метод автоматизированного отбора скважин для применения сшитых полимерных систем (СПС) на основе анализа данных трассерных исследований, выполненных на нефтегазоконденсатном месторождении X. В условиях ограниченного объема размеченных данных разработана двухэтапная модель классификации с использованием алгоритма CatBoost. На первом этапе сформирована интерпретируемая балльная система, основанная на результатах SHAP-анализа многоклассовой модели для сокращения неопределенных случаев. На втором этапе построена модель бинарной классификации, которая также сопровождается интерпретацией результатов с применением SHAP. Предложенный подход позволяет повысить точность и интерпретируемость решений по отбору скважин-кандидатов для проведения технологии СПС, снизить трудоемкость анализа данных и обеспечить инженерную обоснованность результатов.

## **Abstract**

This paper presents a method for automated well candidate selection for the application of cross-linked polymer systems (CPS) based on the analysis of tracer test data obtained from the X oil and gas condensate field. In the context of limited labeled data, a two-stage classification framework is developed using the CatBoost algorithm. At the first stage, an interpretable scoring system is constructed based on SHAP analysis of a multiclass classification model to reduce the number of uncertain cases. At the second stage, a binary classification model is built, also accompanied by SHAP-based interpretation of the results. The proposed approach improves the accuracy and interpretability of decisions related to CPS deployment, reduces the labor intensity of data analysis, and ensures engineering soundness of the outcomes.

**Ключевые** слова: трассерные исследования, сшитые полимерные системы, скважины – кандидаты, отбор скважин, машинное обучение, заводнение, бинарная классификация.

**Keywords:** tracer tests, cross-linked polymer systems, candidate wells, well selection, machine learning, waterflooding, binary classification.

## **1. Введение**

Современная нефтегазовая отрасль сталкивается с необходимостью увеличения нефтеотдачи, особенно в условиях выработанных пластов, характеризующихся резкой неоднородностью и сложным геологическим строением. Одним из перспективных направлений является применение сшитых полимерных систем (СПС), позволяющих селективно блокировать высокопроницаемые интервалы и перераспределять потоки закачиваемого агента в сторону малодренируемых зон [1]. Однако эффективность внедрения данной технологии во многом зависит от правильного выбора скважин-

кандидатов, что требует надежной и объективной оценки фильтрационной связи между нагнетательными и добывающими скважинами.

Трассерные исследования традиционно используются в нефтегазовой отрасли как инструмент оценки фильтрационных каналов, выявления направлений движения закачиваемого агента и диагностики межскважинных связей. Цель трассерных испытаний – получить количественные данные о межскважинной фильтрации: скорости и направления движения закачиваемого агента, наличии гидродинамической связи, объемах охвата заводнением и пр. С их помощью определяют, какая часть закачиваемой воды достигает конкретных добывающих скважин, за какое время и в каких объемах. В частности, анализ динамики появления трассера в продукции добывающих скважин позволяет выявлять зоны быстрого прорыва, связанные с существованием высокопроводящих каналов или трещин, что ведет к неравномерному распределению закачки и росту обводненности [2]. Таким образом, эти исследования позволяют с высокой точностью определить наличие фильтрационного канала, направление и скорость движения флюида, а также интенсивность связи между скважинами.

В настоящее время интерпретация трассерных испытаний преимущественно осуществляется вручную. Специалисты анализируют кривые «концентрация трассера – время», проводят качественный анализ параметров прорыва (время появления трассера, масса закачиваемого реагента, доля извлечения и т.д.) и на их основе принимают решение об эффективности фильтрационной связи [3]. Такой подход является общепринятым, но при этом обладает рядом недостатков и ограничений: он сильно зависит от квалификации инженера, чувствителен к субъективному восприятию и крайне трудоемок при работе с большим массивом данных. В этой связи актуальным становится применение современных алгоритмов машинного обучения, способных автоматизировать обработку данных трассерных исследований, выявлять скрытые закономерности и, что особенно важно, обеспечивать воспроизводимость и интерпретируемость принимаемых решений.

## **2. Описание и подготовка исходных данных**

Для построения модели машинного обучения были использованы результаты трассерных исследований, выполненных на нефтегазоконденсатном месторождении X. В качестве исходной информации использовались производственные параметры закачки, геолого-физические и гидродинамические характеристики, рассчитанные по результатам движения трассера от нагнетательных к контрольным добывающим скважинам. Все данные были собраны из единого отчета по трассерным исследованиям и структурированы в семь различных таблиц, каждая из которых отражала определенную категорию инженерных параметров:

- 1) *df\_params* – параметры добывающих скважин, включая дебит жидкости, нефти и уровень обводненности;
- 2) *df\_output* – информация о времени выхода индикатора на поверхность;
- 3) *df\_volume* – объемно-фильтрационные характеристики проницаемой системы;
- 4) *df\_speed* – скорости фильтрации в межскважинном пространстве;
- 5) *df\_perm* – проницаемость коллекторов по минимуму, максимуму и приведённому среднему значению;
- 6) *df\_area* – геометрические характеристики, включая площадь фильтрационного канала;
- 7) *df\_impact* – информация об обводненности скважин и степени влияния нагнетательной скважины на обводненность контрольной добывающей скважины.

Все данные были предварительно подготовлены и приведены к единому формату с использованием средств Power Query: отформатированы числовые поля, устраниены ошибки типов и унифицированы наименования столбцов. Дальнейший анализ был проведен в Jupyter Notebook с использованием библиотек pandas, matplotlib, seaborn, scikit-learn, shap, catboost.

Все таблицы были объединены в единый датафрейм *df\_merged* на основе пары ключевых идентификаторов: *Number\_expl\_well* (номер добывающей скважины) и *Id\_rump\_well* (номер нагнетательной скважины). Далее проведено переименование столбцов удобства обращения в процессе анализа: все столбцы были переведены на английский язык с сохранением физического смысла (например, *Обводненность, %* → *watercut*, *Масса вынесенного трассера* → *tracer\_mass\_out* и т.д.). Повторяющиеся или избыточные признаки (например, *watercut\_dec*) были удалены. В результате получен единый структурированный датафрейм *df*, содержащий полный набор исходных признаков, необходимых для построения модели.

На следующем этапе был проведен первичный разведывательный анализ данных (EDA), включающий обзор статистик распределения (метод *describe()*) и визуальную проверку выбросов и пропущенных значений. Диапазоны значений по ключевым признакам – дебитам жидкости, трассерным массам, фильтрационным скоростям и объемам породы – оказались широкими, что отражает высокую геологическую и технологическую неоднородность условий. Также анализ позволил выявить пропуск по признаку *hours* (время появления трассера). Ввиду его единичности наблюдение было удалено. Таким образом, итоговый объем данных составил 550 строк, среди которых 194 содержали заполненную целевую переменную *effect*, используемую для обучения модели.

### **3. Формирование логики принятия решений: балльная система и категориальные признаки**

После первичной обработки данных и проведения разведывательного анализа стало очевидно, что исходная экспертная разметка целевой переменной *effect*, отражающая эффективность проведения мероприятий, содержит значительное число пропущенных значений (NaN). Их присутствие ограничивает возможности обучения модели, а ручная доразметка по всем наблюдениям требует значительных трудозатрат и несет риск субъективности в принятии решений.

Для решения данной проблемы был предложен подход, имитирующий экспертную логику, применяемую при интерпретации трассерных исследований. Центральным элементом подхода стала балльная система, позволяющая сократить пропущенные значения путем количественной оценки эффективности по совокупности инженерных признаков.

Для реализации этой системы были сформированы следующие категориальные признаки:

- *breakthrough\_time\_cat* – категория по времени прорыва трассера (1 – короткое время – до 5 суток; 0 – среднее время прорыва – от 5 до 25 суток; -1 – долгое время);
- *tracer\_output\_cat* – категория по массе масса вынесенного трассера (1 – выше среднего, 0 – ниже среднего);
- *tracer\_perct\_cat* – категория по доле вынесенного трассера относительно отобранный массы (1 – выше среднего; 0 – ниже среднего);
- *cat\_impact* – влияние нагнетательной скважины на обводненность добывающей скважины (1 – значительное влияние; 0 – незначительное влияние);
- *cat\_volume* – объем фильтрующей среды (1 при объеме > 60 м<sup>3</sup>; 0 при объеме < 60 м<sup>3</sup>);
- *cat\_watercut* – обводненность (1 – при обводненности более 50%; 0 при обводненности менее 50 %).

Далее для каждой скважины рассчитывалась сводная переменная *effect\_score* как сумма значений этих признаков. На основе итогового балла формировалась переменная *effect\_calc*, принимающая значения:

- 1 – при высоком суммарном балле (эффективная);
- -1 – при низком балле (неэффективная);
- 0 – при промежуточных значениях (неопределенная зона, спорная скважина).

Пороговые значения подобраны таким образом, чтобы не противоречить уже имеющимся экспертным меткам. В результате удалось существенно сократить долю пропущенных значений, получив следующее распределение:

- effect = -1: 297 наблюдений;
- effect = 1: 85 наблюдений;
- effect = 0: 7 наблюдений;
- NaN: 161 наблюдение.

Несмотря на достигнутый прогресс, часть скважин по-прежнему не могла быть отнесена ни к одному классу с достаточной уверенностью. Это послужило основанием для разделения общей задачи на две подзадачи, направленные на дальнейшее уточнение разметки и повышение точности классификации:

- 1) Формирование ранжированной системы признаков и уточнение балльной логики;
- 2) Бинарная классификация на уверенно размеченных данных.

#### **4. Формирование рейтинговой системы и уточнение балльной логики**

Данный этап направлен на выявление признаков, которые оказывают наибольшее влияние на результат классификации. Его основной целью стало построение рейтинговой системы, позволяющей количественно оценить значимость каждого признака и тем самым обосновать обновление логики принятия решений в рамках балльного подхода.

Для этого была обучена модель многоклассовой классификации на подмножестве данных с известной целевой переменной *effect*, принимающей значения: 1, 0, -1. Все пропущенные значения предварительно были переведены в категорию 0, поскольку они отражают неопределенные случаи и по своей природе аналогичны уже имеющимся наблюдениям с классом 0. Также перед построением модели была проведена финальная проверка на наличие высоко коррелирующих признаков. Для этого была построена тепловая карта корреляций и сформирован список признаков с коэффициентом корреляции  $\geq 0.8$ . Анализ показал наличие нескольких пар признаков с высоким уровнем корреляции, что потенциально может негативно сказаться на устойчивости и интерпретируемости модели (рис.1).

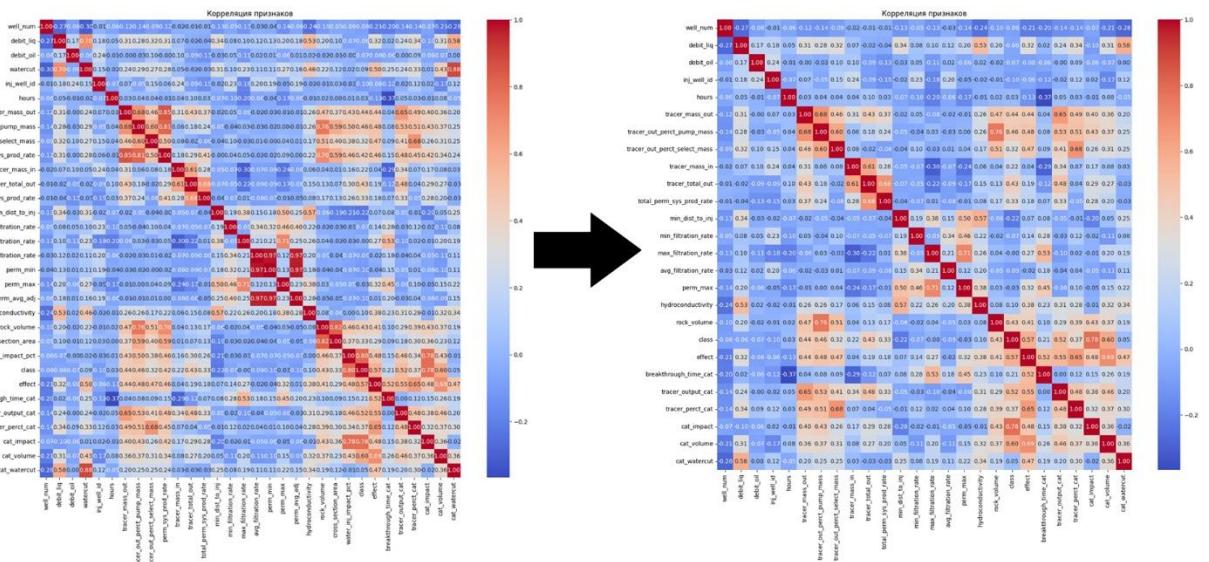


Рисунок 1 - Исходная и обновленная матрицы корреляций

В связи с этим были исключены следующие признаки:

- *perm\_sys\_prod\_rate* – коррелирует с *perm\_avg\_adj*;
- *perm\_min*, *perm\_avg\_adj* – избыточны при наличии *perm\_max*;
- *cross\_section\_area* – дублирует влияние *rock\_volume*;
- *watercut*, *water\_inj\_impact\_pct* – пересекаются по смыслу с *cat\_watercut*.

С учетом наличия как числовых, так и категориальных признаков, а также ограниченного объема выборки и необходимости интерпретируемости, в качестве основной модели был выбран CatBoostClassifier.

CatBoost – это градиентный бустинг на решающих деревьях, разработанный для эффективной работы с категориальными переменными и устойчивости к переобучению на малых выборках [4]. Его преимуществами являются:

- автоматическая обработка категориальных признаков без необходимости one-hot-кодирования;
- устойчивость к несбалансированным выборкам и шумам;
- встроенные методы интерпретации модели (feature importance, SHAP-значения).

Классификация велась по трем уровням целевой переменной (-1, 0, 1), что позволило учитывать не только уверенно эффективные и неэффективные скважины, но и переходные случаи. Обучение проводилось на стандартных параметрах CatBoost, без гиперпараметрической оптимизации. Для построения модели исходный датасет был разделен на обучающую и тестовую выборки в пропорции 80/20 с сохранением баланса классов. Категориальные признаки были переданы в модель через параметр *cat\_features*.

После обучения модель была применена к тестовой выборке, и результаты классификации были оценены с использованием стандартного отчета *classification\_report*

(табл.1) библиотеки Scikit-learn, содержащего значения *accuracy*, *precision*, *recall* и *F1-score* по каждому классу (-1, 0, 1).

При ограниченном количестве эффективных скважин и необходимости точной их идентификации оптимальной метрикой оценки модели является F1-score, объединяющая полноту и точность. Применение данной метрики позволяет повысить вероятность обнаружения перспективных кандидатов при одновременном снижении количества ошибочных рекомендаций.

Таблица 1 - Результаты классификации на тестовой выборке

Class	Precision	Recall	F1-score	Support
-1.0	0,97	0,98	0,97	59
0.0	0,97	0,91	0,94	34
1.0	0,94	1	0,97	17
accuracy			0,96	110
macro avg	0,96	0,96	0,96	110
weighted avg	0,96	0,96	0,96	110

Уже на этом этапе модель продемонстрировала высокую устойчивость и точность классификации по всем классам, включая определение редких случаев (effect = 1). Это подтверждается также построением матрицы ошибок (рис.2), визуализирующей точность классификации по каждому классу.

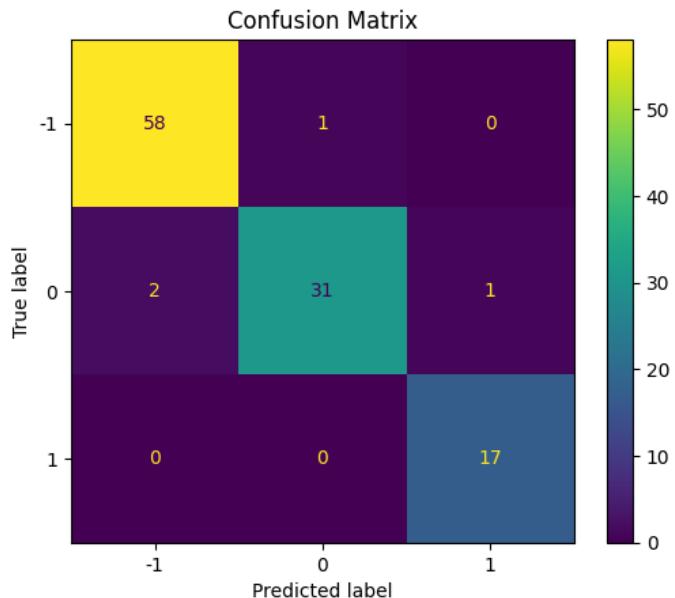


Рисунок 2 - Матрица ошибок

Дополнительно была проведена оценка важности признаков средствами CatBoost (рис. 3). Среди наиболее значимых оказались:

- *breakthrough\_time\_cat* – категория времени прорыва;
- *tracer\_out\_perct\_select\_mass* – выход индикатора в % отобранный массы;

- *rock\_volume* – объем породы;
- *cat\_watercut* – обводненность.

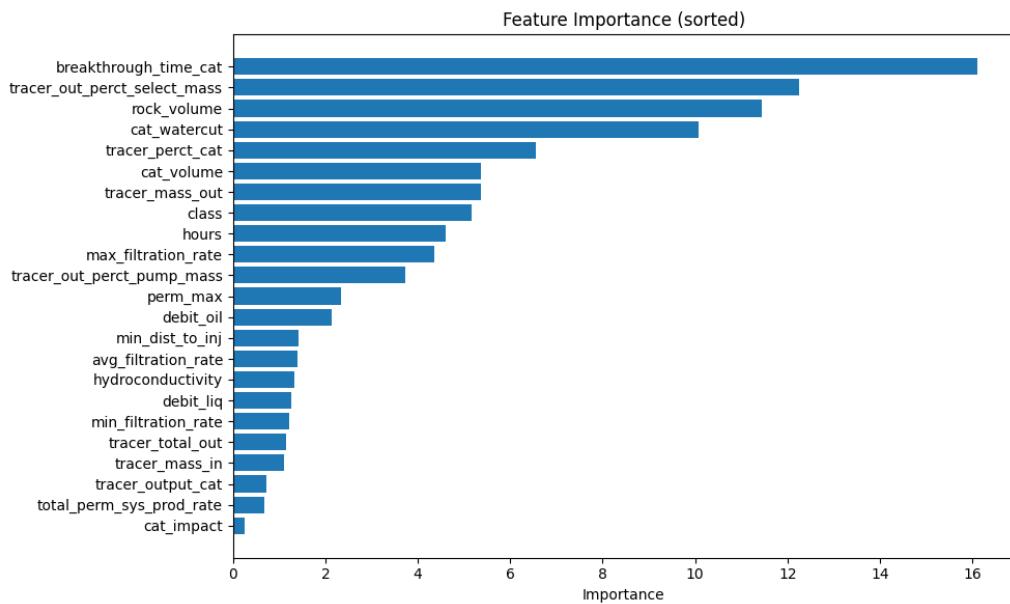


Рисунок 3 - Важность признаков

Однако важно отметить, что данный метод оценивает преимущественно частоту и эффективность использования признака в структуре модели и не отражает его влияние на конкретное предсказание. В связи с этим на следующем этапе был применен более гибкий и интерпретируемый подход – SHAP-анализ, основанный на теории кооперативных игр.

SHAP (SHapley Additive exPlanations) позволяет количественно определить вклад каждого признака в предсказание модели для конкретного объекта. Метод учитывает не только индивидуальные значения признаков, но и их взаимодействие с другими переменными, что особенно актуально в случае сложных зависимостей и мультиклассовой классификации [5].

Для оценки влияния признаков именно на предсказания эффективных/неэффективных скважин, что представляет наибольший интерес в практическом применении модели, был построен beeswarm-график для каждого соответствующего класса (рис. 4). Beeswarm – одна из самых наглядных и мощных визуализаций SHAP. Каждая точка на графике соответствует одному наблюдению, то есть конкретной скважине из набора данных. Цвет точки отражает значение признака: синий – низкое, а красный – высокое. Горизонтальное положение точки показывает направление и силу влияния данного признака на итоговое решение модели – смещение вправо означает увеличение вероятности положительного исхода, тогда как смещение влево указывает на обратный эффект. Признаки упорядочены на оси Y по убыванию их важности, что позволяет быстро оценить, какие параметры оказывают наибольшее влияние на результат.

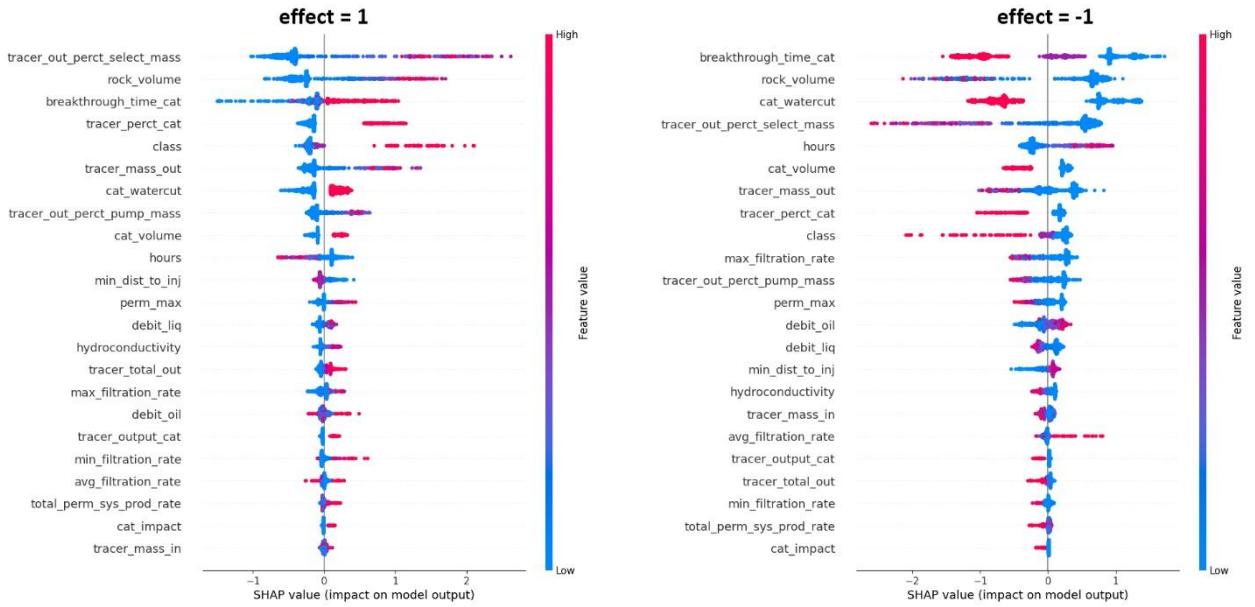


Рисунок 4 - График beeswarm для оценки влияния каждого признака на результат целевой переменной effect = 1(слева) и effect = -1 (справа)

Как видно, разные признаки – в зависимости от класса по-разному влияют на результат, однако в целом оба графика демонстрируют одинаковую логику поведения признаков. И теперь на основании полученных графиков можно выделить веса для наиболее значимых признаков и вывести обновленную формулу расчета итогового балла (*effect\_score\_weighted*):

$$\begin{aligned} \text{effect\_score\_weighted} = & \text{breakthrough\_time\_cat} \cdot 0.30 + \text{class} \cdot 0.19 + \\ & + \text{cat\_watercut} \cdot 0.60 + \text{tracer\_perct\_cat} \cdot 0.30 + \text{cat\_volume} \cdot 0.33 \end{aligned} \quad (1)$$

По значениям итогового балла были установлены пороговые границы:

- score > 1.09 – уверенно эффективная пара скважин (effect = 1);
- score < 0.9 – уверенно неэффективная пара скважин (effect = -1);
- промежуточные значения интерпретировались как effect = 0.

Дополнительно были исключены признаки с весом менее 0.05 (например, *min\_filtration\_rate*, *tracer\_mass\_in*) как шумовые и неустойчивые. Перед проверкой модели вновь была проведена проверка признаков на мультиколлинеарность, после чего получилась финальная матрица корреляции признаков (рис. 5).

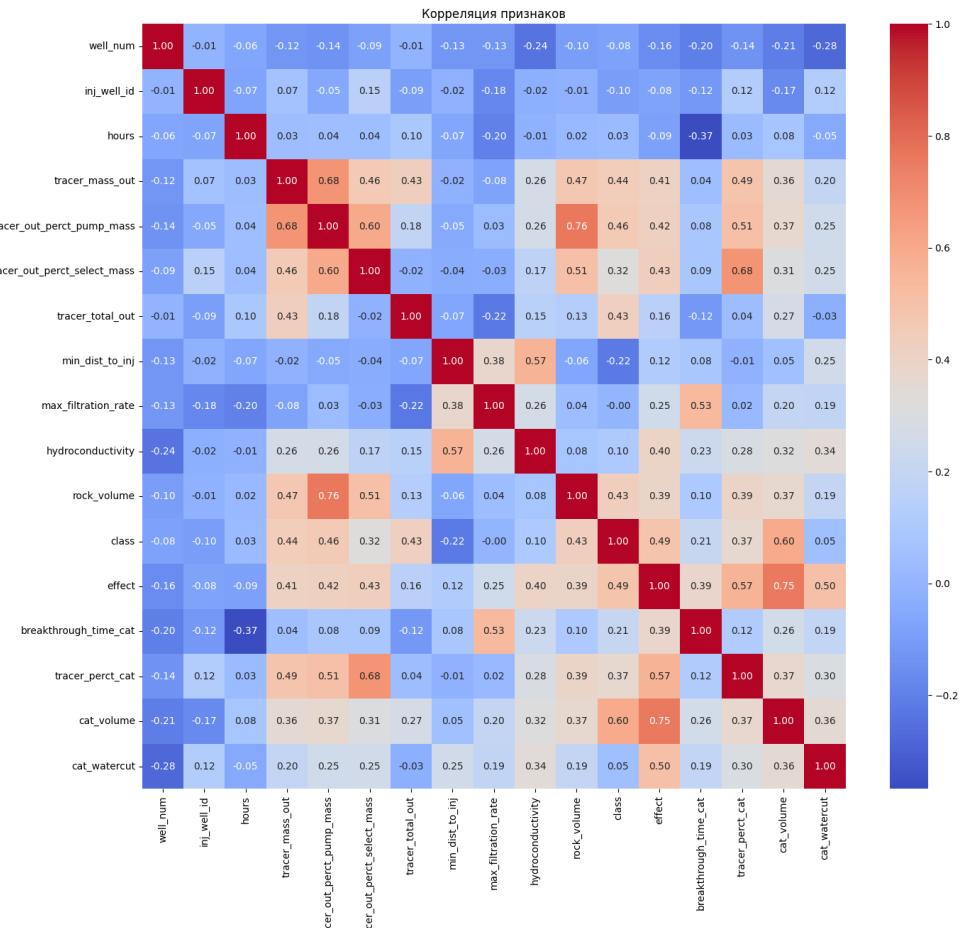


Рисунок 5 - Матрица корреляции конечного набора признаков

Применение уточненной логики позволило значительно сократить количество наблюдений с  $effect = 0$ . В результате только 35 скважин остались в зоне неопределенности и получили статус тестовой выборки, для которой предстояло применить финальную модель бинарной классификации (рис. 6).

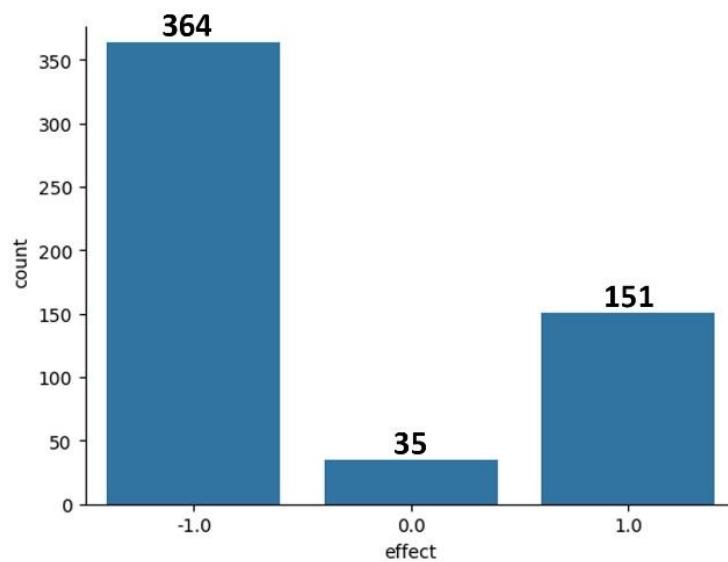


Рисунок 6 - Распределение классов

## 5. Обучение финальной бинарной модели и предсказание по спорной группе скважин

Для построения бинарной модели были отобраны только те наблюдения, где целевая переменная effect имела значения -1 или 1. Это обеспечило формирование чистой обучающей выборки без спорных случаев. В качестве базового алгоритма вновь использовалась модель CatBoost, что позволило обеспечить работу с категориальными признаками без необходимости их дополнительного кодирования, а также сохранить интерпретируемость модели. В результате обучения модель показала отличные результаты на валидационной выборке (табл. 2).

Таблица 2 - Результаты классификации на валидационной выборке

Class	Precision	Recall	F1-score	Support
-1.0	0,99	1	0,99	73
1.0	1	0,97	0,98	30
accuracy			0,99	103
macro avg	0,99	0,98	0,99	103
weighted avg	0,99	0,99	0,99	103

Особенно важно подчеркнуть, что модель не допустила ни одного случая, при котором неэффективная скважина была бы ошибочно отнесена к эффективным (рис.7). Это свидетельствует о высокой надежности модели с точки зрения инженерной и экономической целесообразности, поскольку такой тип ошибки может привести к необоснованным затратам на внедрение СПС в неподходящую зону.

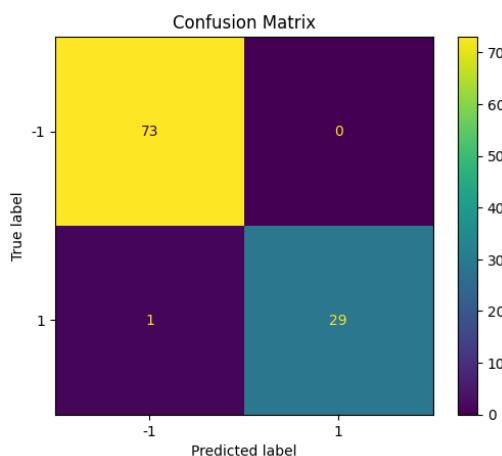


Рисунок 7 - Матрица ошибок для бинарной классификации

В условиях задач заводнения наиболее критичной является именно ошибка первого рода (false positive), при которой неэффективная скважина ошибочно классифицируется как эффективная. Ошибки второго рода (false negative), при которой происходит пропуск действительно эффективной скважины, менее опасны и могут быть компенсированы дополнительным инженерным анализом.

После успешной проверки работы модели на валидационной выборке она была применена к оставшейся тестовой группе наблюдений, для которых сохранялась неопределенность (табл.3). Для повышения интерпретируемости результатов модели была введена шкала уверенности на основе вероятности классификации:

- $\text{confidence} > 0.85$  – уверенно эффективная или уверенно неэффективная;
- $\text{confidence} \leq 0.85$  – неуверенная, рекомендована к дополнительному рассмотрению.

Таблица 3 - Результаты работы модели на тестовой выборке

№ доб.скв.	№ нагн.скв.	Предсказанный effect	Уверенность	confidence
304	361	1.0	0.936259	уверенно эффективная
360	361	-1.0	0.019019	уверенно неэффективная
387	361	1.0	0.919575	уверенно эффективная
391	361	-1.0	0.057755	уверенно неэффективная
1251	361	-1.0	0.066365	уверенно неэффективная
1279	361	-1.0	0.053675	уверенно неэффективная
1307	361	-1.0	0.206733	неуверенная
445	333	1.0	0.946764	уверенно эффективная
4363	306	-1.0	0.009444	уверенно неэффективная
58	362	-1.0	0.431739	неуверенная
279	362	-1.0	0.015916	уверенно неэффективная
304	362	1.0	0.939023	уверенно эффективная
359	362	1.0	0.945732	уверенно эффективная
387	362	1.0	0.934144	уверенно эффективная
445	362	1.0	0.932473	уверенно эффективная
1333	362	1.0	0.994804	уверенно эффективная
58	389	1.0	0.984274	уверенно эффективная
252	389	1.0	0.940136	уверенно эффективная
445	389	1.0	0.852075	уверенно эффективная
360	280	1.0	0.997374	уверенно эффективная
359	307	1.0	0.896394	уверенно эффективная
472	532	1.0	0.518633	неуверенная
1439	532	-1.0	0.481168	неуверенная
501	1471	-1.0	0.039940	уверенно неэффективная
563	1471	-1.0	0.001815	уверенно неэффективная
682	683	-1.0	0.020469	уверенно неэффективная
166	185	-1.0	0.257338	неуверенная
165	185	-1.0	0.212058	неуверенная
205	185	-1.0	0.005864	уверенно неэффективная
4166	185	-1.0	0.005449	уверенно неэффективная
445	388	1.0	0.573800	неуверенная
447	388	-1.0	0.287838	неуверенная
506	1600	1.0	0.984667	уверенно эффективная
1691	1600	-1.0	0.030778	уверенно неэффективная
1662	1693	1.0	0.572920	неуверенная

Результаты предсказания показали, что 26 из 35 скважин были классифицированы как уверенно эффективные или уверенно неэффективные, а остальные 9 наблюдений имели значения вероятности в зоне неопределенности и были оставлены как спорные для дальнейшего инженерного анализа.

Для повышения доверия к результатам и получения прозрачной логики принятия решений были применены два метода интерпретации: встроенная метрика feature importance и SHAP - анализ (рис. 8).

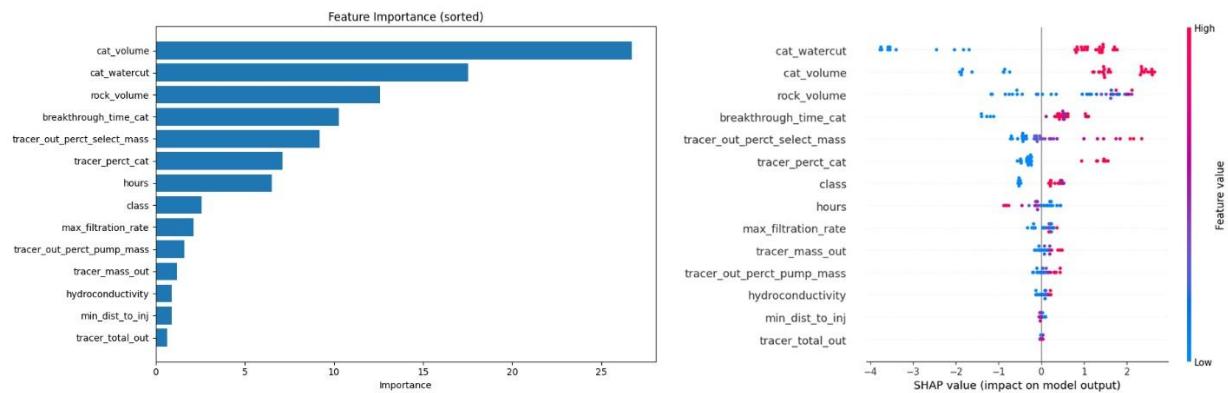


Рисунок 8 - Интерпретация результатов модели на тестовой выборке

Их совместное использование позволило не только выявить ключевые признаки, влияющие на итоговое решение, но и понять, в каком направлении и с какой уверенностью они действуют в отношении класса 1 (эффективная скважина). Согласно полученным графикам, наиболее значимыми оказались:

- 1) *cat\_volume* – категориальная оценка объема фильтрационной среды. Логично, что большие объемы способствуют более равномерному распределению закачиваемой жидкости и являются положительным фактором при выборе скважины;
- 2) *cat\_watercut* – категориальная оценка обводненности. Высокая обводненность, как правило, свидетельствует о гидродинамической связи с нагнетательной скважиной;
- 3) *rock\_volume* – количественная характеристика объема породы, дополняющая категориальную;
- 4) *breakthrough\_time\_cat* – категориальная оценка времени выхода индикатора. Ранний прорыв трассера указывает на наличие зон быстрой фильтрации.

Важно отметить, что по мере перехода от общей к бинарной модели наблюдается усиление роли категориальных признаков, таких как *cat\_volume* и *cat\_watercut*, что указывает на высокую интерпретируемость и доверие к финальной модели.

## Заключение

В данной работе предложена методика применения алгоритмов машинного обучения для анализа трассерных исследований с целью выявления скважин-кандидатов,

подходящих для закачки сшитых полимерных систем (СПС). Рассмотренный подход объединяет инженерные принципы, экспертные оценки и интерпретируемое машинное обучение, что делает его практически применимым в производственных условиях.

В рамках исследования были реализованы следующие ключевые задачи:

- 1) Сформирована интегрированная база данных, включающая параметры трассерных исследований, фильтрационные и геометрические характеристики пласта, данные по закачке и отклику трассера, а также оценки влияния нагнетательных скважин;
- 2) Проведена категоризация признаков на основе инженерных критериев, включая обводненность, время отклика трассера, объемы породы, влияние нагнетательной скважины;
- 3) Разработана и реализована балльная система, имитирующая экспертную логику анализа, что позволило дополнить исходную частично размеченную выборку и сформировать трехклассовую целевую переменную: «эффективно», «неэффективно» и «спорно»;
- 4) Обучена модель CatBoost для многоклассовой классификации, продемонстрировавшая высокую точность ( $accuracy \approx 96\%$ ) и устойчивость при определении классов -1 и 1, а неопределенные случаи ( $effect = 0$ ) были обоснованно оставлены в зоне переходных значений;
- 5) Применены методы интерпретации модели, включая SHAP и feature importance, что позволило выявить значимые признаки и количественно оценить их вклад в каждое решение модели. На основе полученных данных была усовершенствована логика принятия решений с учетом весов признаков в балльной системе;
- 6) Построена финальная модель бинарной классификации, обученная только на уверенно размеченных объектах ( $effect = -1$  и  $effect = 1$ ), которая показала точность 99% на валидационной выборке и не допустила ложной классификации неэффективных скважин как эффективных, что имеет критическое значение с точки зрения промышленной надежности и экономической целесообразности;
- 7) Проведено предсказание по тестовой группе «спорных» скважин, в результате чего более половины объектов были уверенно классифицированы, а оставшиеся рекомендованы для дополнительного инженерного анализа.

В перспективе возможны следующие направления развития предложенного подхода:

- учёт временной динамики параметров, включающих изменение дебитов и обводненности во времени;

- интеграция геопространственной информации - координат скважин и структурных карт;
- дополнение категориальных признаков по другим геологическим и гидродинамическим характеристикам;
- автоматическая настройка гиперпараметров с помощью кросс-валидации и Optuna;
- балансировка классов через oversampling-методы и корректировку весов.

Таким образом, разработанная методика обладает высокой прикладной ценностью для нефтегазовой отрасли. Она обеспечивает автоматизированную обработку больших массивов данных, предоставляет интерпретируемую модель, в которой каждое решение имеет логическое объяснение, снижает вероятность ошибок при выборе скважин-кандидатов и может быть интегрирована в существующие информационные системы как модуль экспертной поддержки принятия решений.

### **Источники литературы**

1. Сафаров Ф.Э., Лобанова С.Ю., Елубаев Б.У., Таламанов Н.Е., Чжидзянь С., Исмагилов Т.А., Телин А.Г. Эффективные методы повышения нефтеотдачи пластов на месторождениях с высоковязкой нефтью: технологии циклического гелеполимерного заводнения и ASP-воздействие // Вестник нефтегазовой отрасли Казахстана. - 2021. - Т. 3. - №3. - С. 61-74
2. Dugstad Ø., Viig S., Krognes B., Kleven R., Huseby O. Tracer monitoring of enhanced oil recovery projects [Электронный ресурс] // EPJ Web of Conferences. – 2013. – Т. 50. – С. 02002. – DOI: 10.1051/epjconf/20135002002. – Режим доступа: [https://www.epj-conferences.org/articles/epjconf/pdf/2013/11/epjconf\\_tracer6\\_02002.pdf](https://www.epj-conferences.org/articles/epjconf/pdf/2013/11/epjconf_tracer6_02002.pdf) (дата обращения: 29.04.2025).
3. Трассерные исследования для оценки эффективности заводнения [Электронный ресурс]. – Режим доступа: [https://www.petroleumengineers.ru/sites/default/files/u4395/trassernye\\_issledovaniya\\_2.pdf](https://www.petroleumengineers.ru/sites/default/files/u4395/trassernye_issledovaniya_2.pdf) (дата обращения: 29.04.2025).
4. CatBoost. Основные этапы алгоритма [Электронный ресурс]. – Режим доступа: <https://catboost.ai/docs/en/concepts/algorithm-main-stages> (дата обращения: 28.04.2025).
5. Чернобровов Д. Интерпретируй это: метод SHAP в Data Science [Электронный ресурс]. – Режим доступа: <https://chernobrovov.ru/articles/interpretiruj-eto-metod-shap-v-data-science.html> (дата обращения: 29.04.2025).