**Оценка проницаемости пород-коллекторов с помощью методов машинного обучения**

**Антропов К.А., *Фролова М.А***

студент, *студент*

Московский государственный университет имени М.В.Ломоносова,физический факультет, Москва, РоссияE–mail: *aantroaleksuu@gmail.com, frolov-tf@mail.ru*

При моделировании движения флюидов в пористой среде важным параметром является проницаемость. Традиционно для её оценки используется связь с пористостью, однако явной зависимости между этими параметрами не существует. Тем не менее на практике часто удается установить полезные корреляционные связи, которые могут быть использованы для анализа. Цель данного исследования — оценить эффективность методов машинного обучения (ML) в прогнозировании проницаемости пород-коллекторов на основе данных о пористости и другой известной информации о месторождениях, чтобы выявить наиболее важные признаки, влияющие на прогнозирование фильтрационных свойств пород-коллекторов.

В настоящей работе использовалась база данных, содержащая наборы параметров четырех реальных нефтяных месторождений. Для анализа данных и установления связи пористости и проницаемости в этом исследовании использовались следующие ML-модели: линейная регрессия (LR), случайный лес (RF), CatBoost [6,7].

 Линейная регрессия была выбрана, так как на практике для установления связи между пористостью и проницаемостью часто используется линейная зависимость в полулогарифмическом масштабе [1].

Второй метод (RF) представляет собой ансамблевый алгоритм машинного обучения, который объединяет множество решающих деревьев для увеличения точности регрессионных предсказаний [2]. В процессе обучения каждое дерево строится на случайно выбранной части данных с произвольным подмножеством признаков, что снижает вероятность переобучения и делает модель более устойчивой. Окончательный результат рассчитывается как среднее значение прогнозов всех деревьев.

Градиентные методы считаются одними из наиболее эффективных в анализе данных, так как работают по принципу последовательного уменьшения ошибок предыдущих моделей. Одним из наиболее известных алгоритмов этого типа является CatBoost [5], который оптимизирован для работы с категориальными признаками и отличается высокой точностью. Модель обучается на основе градиентного спуска, где каждая следующая итерация исправляет ошибки предыдущей, что значительно улучшает качество предсказаний.

Для оценки качества моделей использовался коэффициент детерминации *R*2, рассчитанный с помощью кросс-валидации.

Также были рассмотрены некоторые теоретические модели связи проницаемости и пористости горных пород, полученные для различных упрощенных представлений о структуре пористой среды [3].

Результаты показали, что ML-модели, особенно CatBoost, превосходят теоретические зависимости по точности прогноза (табл. 1). Например, для месторождения 'D' CatBoost достиг R2 = 0,79, в то время как лучшая теоретическая модель показала R2 = 0,69. Однако для некоторых месторождений, таких как 'B' и 'C', достичь высоких значений R2 не удалось. Добавление дополнительных признаков позволило улучшить результаты. Например, для месторождения 'B' использование дополнительных данных повысило *R*2 с 0,174 до 0,659 при использовании CatBoost.

Для различных месторождений признаки, вносящие наибольший вклад в улучшение качества градиентного бустинга, отличаются. Тем не менее в большинстве случаев, помимо пористости, на прогнозирование проницаемости наибольшее влияние оказывал тип порового пространства. Информация об остаточной водонасыщенности образцов при различных капиллярных давлениях также является полезным фактором.

***Таблица 1***. Результат применения теоретических уравнений и методов машинного обучения для четырех реальных нефтяных месторождений.

|  |  |
| --- | --- |
|  | R2 |
| Месторождение |
| ‘A’ | ‘B’ | ‘C’ | ‘D’ |
| Физическиемодели | model 1 | 0,733 | 0,165 | 0,184 | 0,628 |
| model 1\* | 0,773 | 0,133 | 0 | 0,692 |
| model 2 | 0,384 | 0 | 0 | 0,608 |
| model 3 | 0,300 | 0 | 0 | 0,566 |
| model 4 | 0,369 | 0,082 | 0,181 | 0,293 |
|  | model 5\*\*\* | 0,719 | 0,152 | 0,290 | 0,683 |
| Моделимашинногообучения | Linear Regression | 0,743 | 0,143 | 0,184 | 0,746 |
| CatBoost | 0,852 | 0,174 | 0,353 | 0,789 |
| CatBoost\*\* | 0,856 | 0,659 | 0,421 | 0,831 |
| Random Forest\*\* | 0,874 | 0,688 | 0,457 | 0,848 |

\* модифицированная модель; \*\* модели машинного обучения, в которых помимо признаков 'пористость' и 'проницаемость' использовались и другие доступные данные; \*\*\* model\_5 представляет из себя комбинацию физической модели и использования оптимизационных методов, таких как алгоритм обобщенного приведенного градиента или эволюционный поиск (методы ML)

Таким образом, были рассмотрены некоторые теоретические зависимости, устанавливающие связь пористость-проницаемость, а также их сравнение с результатами применения алгоритмов машинного обучения. В большинстве случаев ML-модели позволяют достичь большего значения коэффициента детерминации по сравнению с теоретическими зависимостями. Самым «сильным» алгоритмом оказался CatBoost. Не для всех месторождений удаётся достичь высокого значения R2. При этом, если добавить в анализ месторождения некоторые дополнительные данные, удается повысить метрику оценки качества модели, иногда в несколько раз. В этом случае наибольшие показатели дал алгоритм RF.

**Литература**

1. Рекомендации к методике построения геологических моделей при подсчете запасов углеводородного сырья. Москва: ФБУ «ГКЗ», 2014.
2. Breiman L. Random forests //Machine learning. – 2001. Т. 45. С. 5-32.
3. Gary Mavko, Tapan Mukerji, Jack Dvorkin. The Rock Physics Handbook // Cambridge University Press, 2020, P. 525-542, 978-1-108-42026-6.
4. Hilary L. S. The historical development of the Gauss linear model //Biometrika. 1967, Т. 54. – №. 1/2. С. 1-24.
5. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features //Advances in neural information processing systems. 2018. Т. 31.
6. Sadeqi-Arani Z., Kadkhodaie A. A bibliometric analysis of the application of machine learning methods in the petroleum industry //Results in Engineering. 2023, Т. 20. С. 101518.
7. Watt J., Borhani R., Katsaggelos A. K. Machine learning refined: Foundations, algorithms, and applications. – Cambridge University Press, 2020.