

$$Y_x = 1,53 + (-0,994) \frac{0,087}{1,22} (X_i - 4,85) = 1,875 + 0,071X_i$$

Применяя исходные условия обозначений, уравнение (2) примет вид (3):

$$\sum Q_n = 1,875 + 0,071\eta_n \quad (3)$$

По уравнению (3), результатам расчётов и исходным данным мы строим кривые фактических и теоретических значений накопленной добычи от содержания нефти (рис. 1).

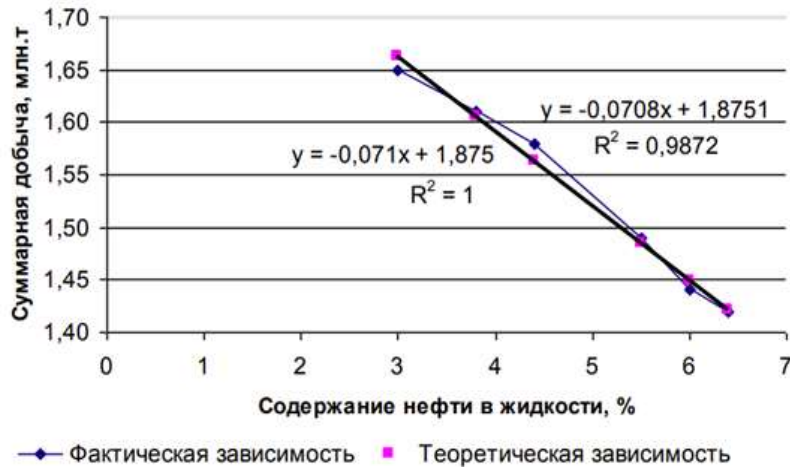


Рис. 1. Кривые фактических и теоретических значений накопленной добычи от содержания нефти

При  $\eta_n = 0$  получаем потенциальные извлекаемые запасы нефти, которые будут равны свободному члену уравнения, то есть при  $\eta_n = 0$ ;  $Q_n = 1,88$  млн.т. По обоснованному конечному значению параметра ( $\eta_n$ ) находим промышленные извлекаемые запасы нефти ( $\sum Q_n$ ). В этом случае остаточные извлекаемые запасы нефти определяются по формуле:

$$Q_{н.ост.} = Q_n - Q_{н.доб.} \quad (4)$$

$Q_{н.ост.}$  – остаточные извлекаемые запасы нефти;  
 $Q_n$  – промышленные извлекаемые запасы нефти;  
 $Q_{н.доб.}$  – общая добыча нефти на дату подсчёта.

#### Литература

1. Мирзаджанзаде А. Х. Технология и техника добычи нефти [Текст] / А. Х. Мирзаджанзаде. – М.: «Недра», 1986. – 382 с.
2. Митропольский А. К. Техника статистических вычислений [Текст] / А. К. Митропольский. – М.: Наука, 1971. – 576 с.
3. Мищенко И. Т. Расчёты в добыче нефти [Текст] / И. Т. Мищенко. – М.: «Недра», 1989. – 245 с.

### РАЗРАБОТКА КАСКАДНОЙ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ ЭФФЕКТА ОТ ГРП

Кондратенко Ф.И., Мурачев А.С.<sup>1</sup>

Научный руководитель А.А. Юркин<sup>2,3</sup>

<sup>1</sup>Санкт-Петербургский политехнический университет Петра Великого,  
г. Санкт-Петербург, Россия

<sup>2</sup>Национальный исследовательский Томский политехнический университет, г. Томск, Россия

<sup>3</sup>ПАО «Славнефть-мегионнефтегаз», г. Мегион, Россия

В настоящее время наиболее распространенным методом интенсификации добычи углеводородов является гидравлический разрыв пласта (ГРП) [1,2], оценка эффекта от которого, как правило, требует привлечения высококвалифицированного персонала и существенные затраты времени для оптимизации дизайна. Недостатками классического подхода к расчету ГРП являются повышенные операционные затраты на подготовительном этапе и невозможность автоматизированного оперативного выполнения оценки потенциала для всего фонда скважин с высокой точностью, даже при наличии исторических данных. В то же время, современное развитие нефтяной отрасли предполагает повышение скорости выполнения операций и степени цифровизации, что должно снизить операционные затраты на подготовительном этапе. В данной статье авторами предлагается автоматизированная методика оценки эффекта от ГРП на основе исторических данных с использованием методов машинного обучения (ML).

Для обучения ML-методов и разработки математического алгоритма применялась выгрузка из базы данных компании Х. Колонки в датасете можно по смыслу разделить на пять групп: параметры работы скважины на момент

остановки, фактические запускные параметры, прогнозные модельные параметры, данные о дизайне ГРП, информация о пласте и его флюидах. Количество записей в датасете 5077. Каждая из записей является отдельным случаем проведения ГРП.

В исходных данных наблюдалось порядка 20-30 % пропущенных и/или некорректно занесенных значений в каждом столбце. Значения, где это было возможно, заполнялись с помощью трех подходов: замена средним арифметическим, замена математическим ожиданием, расчет значения по остальным параметрам с помощью ML-алгоритма. Наиболее качественным результатом оказалось заполнение с помощью ML-модели. Некорректные значения были исправлены при помощи алгоритмов фильтрации данных.

Для улучшения качества алгоритмов в датасет были дополнительно добавлены следующие параметры: данные инклинометрии, характеристика вытеснения нефти (формула Сазонова), данные о работе скважин окружения (средние значения для дебита, обводнённости, приемистости и др.). Также были добавлены нелинейные параметры, такие, как: логарифм остановочной обводнённости, дистанция между забоем и устьем скважины, азимут с забоя на устье скважины и др. После добавления всех параметров была построена тепловая карта корреляций для всех параметров в датасете (рис.1). По данной таблице корреляций были отброшены столбцы, имеющие низкую корреляцию с запускными параметрами скважины, например, такие как высота спуска НКТ, азимут перфорации.

Для обучения датасет был разделен на две выборки: тестовую и обучающую. Размер тестовой выборки-15 % от общего объема датасета. На обучающей выборке были обучены [9] следующие алгоритмы [8]: RandomForest [3], Catboost, XGBoost [5]. В ходе обучения было замечено, что при построении прогноза Catboost и RandomForest опираются на различные признаки, следовательно, предсказания данных моделей независимы. Это позволило объединить и использовать их параллельно, то есть, на основе предсказаний этих двух моделей и входных данных можно выполнить еще один прогноз. При этом, для обучения датасет изменялся, исходя из трех сценариев использования получившегося алгоритма в долгосрочной перспективе: «положительного», наиболее «реального» и «негативного». «Положительный» сценарий подразумевает наличие в датасете всех данных, включая гидродинамическую модель дизайна ГРП и расчетные запускные параметры по математической модели. Наиболее «реальный» сценарий подразумевает отсутствие модельных данных об ожидаемых запускных параметрах скважины, но наличие гидродинамических испытаний скважины. «Негативный» сценарий подразумевает отсутствие всего кроме PVT-параметров, параметров пласта и входных данных о примерном дизайне ГРП.

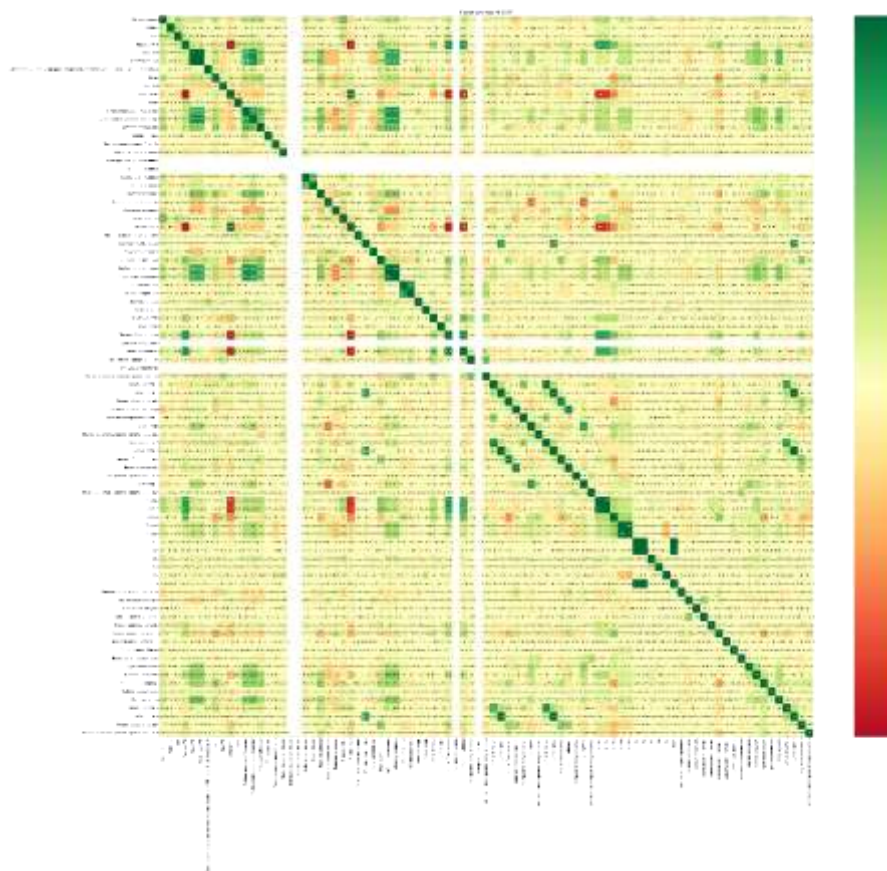


Рис. 4. Тепловая карта корреляций датасета

В ходе работы были также получены зависимости, описывающие определенные внешние факторы: так, например, ключевым параметром для предсказания запускной обводнённости являлся исполнитель (подрядчик) ГРП. На основе этого можно сделать вывод о значительном влиянии используемых химических составов подрядных

организаций на запусковое значение обводненности скважины. Распределение влияния параметров при построении прогноза приведено на рисунке 2:



Рис. 5. Распределение влияния параметров при построении запусковой обводненности

Наиболее точные результаты предсказаний были получены для положительного сценария. Результаты предсказаний для негативного и наиболее реального сценариев применения отличаются в пределах 5-10 %. Все метрики результатов предсказаний моделей приведены в таблице 1, где представлено сравнение прогноза по ГДИ с результатами работы ML-модели для тестовой выборки «реального» сценария:

Таблица 1

Сравнение результатов прогноза

Предсказываемый параметр		Средняя абсолютная ошибка	Доля прогнозов с относительной погрешностью более 15 %	Доля прогнозов с относительной погрешностью более 5 %	$R^2$
ГДИ	Дебит жидкости	18.47 м <sup>3</sup>	0.71	0.65	0.48
	Обводненность	5.57 %	0.71	0.70	0.91
	Дебит нефти	7.1 тонн	0.71	0.65	0.96
Прогноз ML	Дебит жидкости	6.9 м <sup>3</sup>	0.80	0.71	0.81
	Обводненность	0.93 %	0.85	0.81	0.95
	Дебит нефти	2.5 тонн	0.81	0.73	0.99

Разработанная модель позволяет получить результаты, в среднем превосходящие результаты аналитического прогноза. Таким образом, в ходе работы была получена каскадная модель, позволяющая получать расчет запусковых параметров скважины после ГРП с точностью, сопоставимой с аналитическими моделями. Ключевой эффект от использования полученной разработки – снижение времени, так как инференс уже обученной ML-модели производится за крайне небольшое время (до 5 секунд, в зависимости от данных). Также стоит отметить меньшую среднюю абсолютную ошибку, что позволит строить более точный прогноз суммарно по всем кандидатам.

#### Литература

1. Мухина Е. А. Физико-химические методы анализа. – М.: Химия, 1995. – 415 с.
2. Реймерс Н.Ф. Природопользование. – М.: Мысль, 1990. – 637 с.
3. L. Breiman. Random forests. *Maching Learning*, 45(1):5 – 32, Oct. 2001.
4. T. Chen, H. Li, Q. Yang, and Y. Yu. General functionalmatrix factorization using gradient boosting. (ICML'13), раздел 1, с. 436 – 444, 2013.
5. J. Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4): 367 – 378, 2002.
6. J. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5):1189 – 1232, 2001.
7. T. Zhang and R. Johnson. Learning nonlinear functions using regularized greedy forest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5), 2014.
8. G. Ridgeway. Generalized Boosted Models: A guide to the gbm package
9. X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, D. Xin, R. Xin, M. J. Franklin, R. Zadeh, M. Zaharia, and A. Talwalkar. MLlib: Machine learning in apache spark. *Journal of Machine Learning Research*, 17(34):1–7, 2016.