

ПРОГНОЗИРОВАНИЕ ОТКАЗОВ НЕФТЯНЫХ СКВАЖИН С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВАНИИ ДАННЫХ ЭКСПЛУАТАЦИОННОГО ФОНДА СКВАЖИН

Мадаминджанов А.Ю., Нуриева Д.Ф., студенты III курса магистратуры по искусственному интеллекту и цифровым двойникам в топливно-энергетическом комплексе, ФГБОУ ВО «УГНТУ»
Захаров А.В., к.ф.-м.н., доцент, ФГБОУ ВО «УГНТУ»

Аннотация. Статья посвящена использованию машинного обучения для прогнозирования отказов скважин нефтяного фонда.

Ключевые слова: нефтегазовая отрасль, анализ данных, машинное обучение.

Нефтегазовая промышленность по праву многими считается стратегической отраслью нашего государства: это основа развития экономики и обеспечение стабильности страны.

Все в современном мире зависит от топлива:

- Транспортные средства передвигаются благодаря нефтепродуктам;
- Из природного газа изготавливаются полимеры;
- Многие медикаменты производятся из материалов нефтяной и газовой переработки;
- Даже детские игрушки производятся из продукции нефтяной и газовой промышленности.

Все это делает газ и нефть одними из самых полезных и полезных ресурсов не только топливной промышленности, но и остальных сфер. Соответственно в экономике РФ нефтегазовая отрасль играет ключевую роль.

Газовая и нефтяная сферы тесно связаны с другими производствами и экономиками России. Товарооборот внутри государства и за его пределами полон продуктами переработки нефти и газа. Эти ресурсы значительно превышают другие статьи экспорта и импорта. Соответственно энергетический баланс сильно зависит от получаемого природного газа и нефти. Если

выполнять расчет на нефтяной эквивалент, то на нефть и газ приходится примерно 60% всемирного потребления источников энергии.

Наиболее крупными предприятиями нефтегазовой промышленности в России являются:

– ПАО «Лукойл»; ПАО «Роснефть»; ПАО «Газпром»; ПАО «Сургутнефтегаз»[1].

В настоящее время мировая промышленность, включающая в себя нефтегазовую отрасль, входит в эру изменений колоссального масштаба, которую К. Шваб охарактеризовал как «четверную промышленную революцию»[2]. Двигателями этого процесса являются достижения в роботизации, интернета вещей, искусственного интеллекта и в других направлениях, например, нанотехнологиях[3].

Для поддержания ведущего уровня развития введен тренд на цифровизацию.

Одно из важнейших направлений цифровизации являются технологии искусственного интеллекта. Наибольшее развитие получают направления, которым можно отнести технологии компьютерного зрения и естественного языка, роботизированная автоматизация процессов, разработка различных виртуальных помощников и расширение машинного обучения[4]. Исходя из вышеизложенного в рамках очень важно применять новейшие технологии в добыче углеводородов.

Одной из проблем в добыче углеводородов являются аварии и отказы оборудования на нефтяных скважинах. При авариях возможен вред от разливов нефти и их последствия. Как и при авариях и отказах оборудования кроме возможного негативного влияния падает и уровень добычи нефти. Соответственно, прогнозирование отказов, куда входят и аварии, может являться одним из приоритетных и актуальных направлений для использования средств искусственного интеллекта.

Примерами прогнозирования отказов и контроля состояний скважин являются:

1. Совершенствование системы контроля состояния скважинной штанговой глубинной насосной установки на основе нейросетевых технологий[5];

2. ИС «Солярис»[6];

3. Повышение эффективности эксплуатации скважин с помощью инструментов интегрированного моделирования»[7].

В данных примерах прогнозирование проводится под определенный тип установленного оборудования. Мы же в своей работе для обучения моделей прогнозирования отказов скважин используем огромный пласт данных эксплуатационного фонда скважин.

Для анализа данных эксплуатационного фонда скважин были выбраны данные за 1 месяц в одном из предприятий нефтегазовой отрасли.

В ходе анализа данных были подготовлены для следующего использования следующие данные:

– Признак остановки скважины в течении 5 дней; Глубина спуска насоса. После ремонта; Дебит жидкости. Запускной; Обводненность. Запускная; Дебит нефти. Запускная; Дебит жидкости. Только из тех режима; Обводненность. Только из тех режима; Дебит жидкости. Только из тех режима. Текущая наработка, для скважин в работе; Дебит жидкости. Текущий режим; Обводненность. Текущий режим; Дебит нефти. Текущий режим; Газовый фактор. Текущий режим; Дебит газа. Текущий режим; Давление пластовое. Текущий режим.

Построена матрица корреляции данных (Рисунок 1).

Нас интересует признак отказа скважины в течении 5 суток – IS_EXISTS_STOP. Как мы видим из матрицы корреляции все параметры оказывают влияние на данный признак и мы можем далее использовать эти обработанные данные.

Так как мы имеем прогнозируемый параметр, и он имеет классификацию, то соответственно выбрана стратегия обучения с учителем – классификация.

	IS_EXISTS_STOP	PUMP_HSP_AW	LR_WS	W_WS	OR_WS	LR_TR	W_TR	OR_TR	MRP_CURR	LR_TRC_AW	W_TRC_AW	OR_TRC_AW	GF_TRC_AW	GR_TRC_AW	PPL_TRC_AW
IS_EXISTS_STOP	1.000000	0.023000	0.003000	-0.002000	0.009000	-0.005000	-0.007000	-0.002000	-0.018000	-0.011000	-0.008000	-0.008000	0.001000	-0.002000	0.002000
PUMP_HSP_AW	0.023000	1.000000	-0.079000	-0.215000	0.196000	-0.091000	-0.230000	0.161000	-0.285000	-0.095000	-0.223000	0.156000	0.378000	0.266000	0.330000
LR_WS	0.003000	-0.079000	1.000000	0.370000	0.329000	0.950000	0.298000	0.343000	-0.171000	0.949000	0.295000	0.343000	-0.066000	0.170000	0.127000
W_WS	-0.002000	-0.215000	0.370000	1.000000	-0.410000	0.341000	0.793000	-0.306000	0.156000	0.341000	0.797000	-0.301000	-0.140000	-0.246000	0.141000
OR_WS	0.009000	0.196000	0.329000	-0.410000	1.000000	0.321000	-0.330000	0.845000	-0.226000	0.318000	-0.334000	0.838000	0.058000	0.550000	0.038000
LR_TR	-0.005000	-0.091000	0.950000	0.341000	0.321000	1.000000	0.290000	0.380000	-0.172000	0.992000	0.281000	0.378000	-0.071000	0.177000	0.116000
W_TR	-0.007000	-0.230000	0.298000	0.793000	-0.330000	0.290000	1.000000	-0.394000	0.202000	0.290000	0.971000	-0.381000	-0.131000	-0.303000	0.131000
OR_TR	-0.002000	0.161000	0.343000	-0.306000	0.845000	0.380000	-0.394000	1.000000	-0.233000	0.375000	-0.389000	0.978000	0.043000	0.613000	0.025000
MRP_CURR	-0.018000	-0.285000	-0.171000	0.156000	-0.226000	-0.172000	0.202000	-0.233000	1.000000	-0.169000	0.201000	-0.225000	-0.055000	-0.151000	-0.048000
LR_TRC_AW	-0.011000	-0.095000	0.949000	0.341000	0.318000	0.992000	0.290000	0.375000	-0.169000	1.000000	0.282000	0.379000	-0.074000	0.175000	0.113000
W_TRC_AW	-0.008000	-0.223000	0.295000	0.797000	-0.334000	0.281000	0.971000	-0.389000	0.201000	0.282000	1.000000	-0.398000	-0.139000	-0.321000	0.135000
OR_TRC_AW	-0.008000	0.156000	0.343000	-0.301000	0.838000	0.378000	-0.381000	0.978000	-0.225000	0.379000	-0.398000	1.000000	0.041000	0.625000	0.020000
GF_TRC_AW	0.001000	0.378000	-0.066000	-0.140000	0.058000	-0.071000	-0.131000	0.043000	-0.055000	-0.074000	-0.139000	0.041000	1.000000	0.435000	0.099000
GR_TRC_AW	-0.002000	0.266000	0.170000	-0.246000	0.550000	0.177000	-0.303000	0.613000	-0.151000	0.175000	-0.321000	0.625000	0.435000	1.000000	0.081000
PPL_TRC_AW	0.002000	0.330000	0.127000	0.141000	0.038000	0.116000	0.131000	0.025000	-0.048000	0.113000	0.135000	0.020000	0.099000	0.081000	1.000000

Рис. 1. Матрица корреляции данных

Проведена разбивка целевого параметра на классы (Рисунок 2).

```
HSWD['IS_EXISTS_STOP'].value_counts(bins= 2)
(-0.002, 0.5]    76658
(0.5, 1.0]       758
Name: count, dtype: int64
```

Рис. 2. Разбиение на классы «IS_EXISTS_STOP»

Соответственно значения 0 – отсутствие остановки/отказа в следующие 5 суток и 1 – наличие остановки/отказа в следующие 5 суток.

Как мы видим из результата выборка классов оказалась сильно несбалансированной:

- 99% класса отсутствия остановки;
- 1% класса наличия остановки в течении следующих 5 дней.

Соответственно можно сделать вывод, что для обучения моделей необходимо будет проводить семплирование данных.

Для обучения модели будем использовать методы классификации линейную, ближайших соседей, случайного леса. Предварительно проведем сэмплирование данных с помощью методов SMOTE, ADASYN, TomekLinks. Выберем лучший результат и обоснуем его используя различные метрики.

Нас интересует только один класс – отказы. Для удобства данные отобразим в сравнительной таблице методов классификаций (Таблица 1).

Таблица 1 – Сравнение метрик обученных моделей

№ п/п	Метод классификации	Precision	Recall	F1
1	Линейная со SMOTE	0.02	0.53	0.03
2	Линейная с ADASYN	0.02	0.55	0.03
3	Ближайших соседей со SMOTE	0.73	0.86	0.79
4	Ближайших соседей с ADASYN	0.73	0.86	0.79
5	Ближайших соседей с TomekLinks	0.78	0.54	0.64
6	Случайного леса со SMOTE	0.91	0.78	0.84
7	Случайного леса с ADASYN	0.91	0.80	0.85
8	Случайного леса с TomekLinks	0.92	0.76	0.83

Где Precision – точность, которая показывает количество истинно положительных исходов из всего набора положительных меток. Recall – чувствительность, определяет количество истинно положительных среди всех меток класса, которые были определены как «положительный». F1 – среднее гармоническое для получения оценки результатов.

Нас интересует как чувствительность, так и точность модели, соответственно можно брать значение f1. По результатам сравнения нам наиболее подошла классификация методом случайного леса с ADASYN сэмплированием. Результаты метрик отображены на рисунках 3 и 4.

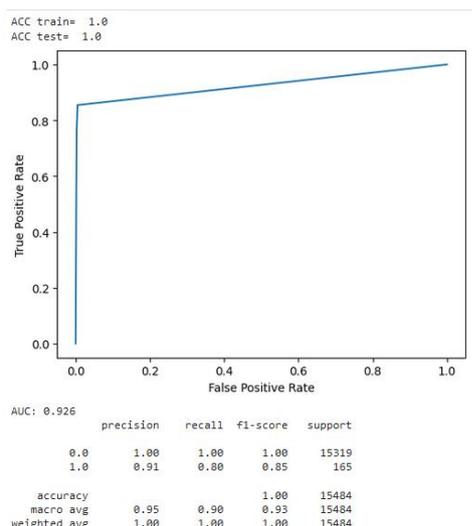


Рисунок 3 – Метрика точности и ROC кривая для классификации методом случайного леса с сэмплированием ADASYN

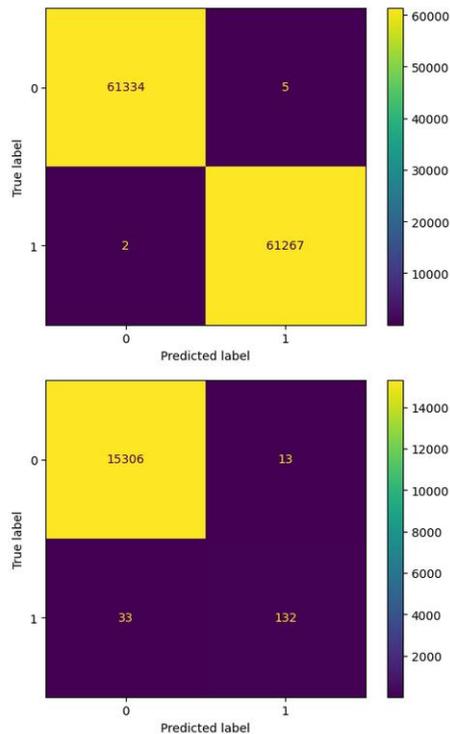


Рисунок 4 – Матрица смежности для классификации методом случайного леса с семплированием ADASYN

Среднее гармоническое для получения оценки результатов нашей модели – 0.85, что является хорошей и достаточной вероятностью прогноза отказов скважин и соответственно с данной моделью мы можем протестировать данную модель в опытной эксплуатации.

При последующем успешном тестировании и дальнейшем внедрении модели этого решения позволит своевременно предсказывать отказы, в том числе и аварии, на скважинах нефтяного фонда. За счет чего получится сократить себестоимость добычи нефти и продлить период разработки. Внедрение моделей машинного обучения является экономически выгодным, так как не требует капитальных затрат на покупку оборудования или химических компонентов. Кроме того, сотрудники получают доступ к инструментам анализа данных, получаемых на месторождениях для предотвращения отказов.

Литература

1. Нефтегазовая промышленность России: роль, тенденции, развитие [Электронный ресурс]//URL:<https://www.neftegaz-expo.ru/ru/articles/neftegazovaya-promyshlennost-rossii/> (дата обращения 13.11.2024).
2. Шваб К. Четвертая промышленная революция. М.: Эксмо, 2023. 208 с..
3. Technological Prerequisites and Humanitarian Consequences of Ubiquitous Computing and Networking / A. Guryanova, E. Khafiyatullina, M. Petinova, V. Frolov, A. Makhovikov // Lecture Notes in Networks and Systems. 2020. Vol. 87. P. 1040–1047.
4. Гурьянов Н. Ю., Гурьянова А. В. Цифровая глобализация в контексте развития цифровой экономики и цифровых технологий // Вестник Московского государственного областного университета. Серия: Философские науки. 2020. № 3. С. 63–69. DOI: 10.18384/2310-7227-2020-3-63-69
5. Латыпов Б.М., Гумерова В.И. Совершенствование системы контроля состояния скважинной штанговой глубинной насосной установки на основе нейросетевых технологий // Нефтегазовое дело. 2022. Т. 20, № 6. С. 155-164. <https://doi.org/10.17122/ngdelo-2022-6-155-164>.
6. Техника и технология нефтехимического и нефтегазового производства: материалы 11-й Междунар. науч.-техн. конф. (Россия, Омск, 24–27 февр. 2021 г.) / М-во образования Ом. обл. [и др.] ;редкол.: В. А. Лихолобов [и др.]. – Омск: Изд-во ОмГТУ,2021. – 292 с.: ил. С. 170-173. ISBN 978-5-8149-3202-0
7. Известия Томского политехнического университета. Инжиниринг георесурсов. 2022. Т. 333. № 8. 109–117 Жигалов Д.Н., Поплыгин В.В. Повышение эффективности эксплуатации скважин с помощью инструментов интегрированного моделирования. <https://doi.org/10.18799/24131830/2022/8/3061>.