Краткое описание

**Прогноз дебита нефти**

Необходимо построить модель для выполнения среднесрочного прогноза добычи нефти на основе ретроспективных данных по объемам дебита нефти и другим параметрам работы скважин.

Задача представлена Департаментом искусственного интеллекта и анализа данных ООО «Газпромнефть – Цифровые Решения»

---------------------------------------------------------------------------------------------------------------------------------------

Подробное описание

**Описание задачи:**

Ваша цель – построить модель для прогнозирования дебита нефти на 90 дней вперед, опираясь на предшествующие значения прогнозируемого показателя и (опционально) других параметров работы скважины.

Дебит нефти – объем ежедневно добываемой из скважины нефти. Является ключевой характеристикой ее работы, определяющей способность скважины генерировать продукт при заданном режиме эксплуатации.

Дебит нефти зависит от многих факторов:

* От этапа жизненного цикла скважины: наиболее высокая добыча нефти наблюдается на начальном этапе эксплуатации скважины, в дальнейшем, по мере исчерпания запасов, объемы дебита нефти снижаются, в то время как доля воды в поступающем из скважины флюиде растет. Нормативный период эксплуатации скважины составляет 15 лет.
* От запасов нефти в пласте, из которого производится добыча. При прочих равных условиях скважины разных месторождений будут показывать разный дебит нефти. Даже скважины в пределах одного месторождения в зависимости от местоположения могут отличаться динамикой показателей.
* От режима работы насосов: изменение режима работы скважины может приводить как к увеличению, так и к уменьшению дебита нефти. Отследить смену режима эксплуатации можно по динамике забойного давления.
* От технического состояния скважины: при возникновении различных дефектов скважины доля воды в общем объеме флюида резко и значительно возрастает, а доля нефти снижается. После проведения ремонтных работ на скважине дебит нефти восстанавливается. На графике это можно идентифицировать по характерным паттернам показателей дебита нефти и воды и производных от них коэффициентов.

Для диагностики состояния скважин используют показатель обводненности, рассчитываемый по формуле:

Qнефти / (Qнефти + Qводы) \* 100%

Также используется водонефтяной фактор или WOR (Water Oil Ratio), рассчитываемый как:

Qводы / Qнефти

Модели, способные с высокой точностью прогнозировать дебит нефти повысят эффективность эксплуатации нефтяных скважин и скорость реагирования на негативные изменения ключевых показателей.

Сложность задачи связана с тем, что на работу скважины оказывает влияние большое количество факторов, которые не всегда возможно учесть и проводить их мониторинг на регулярной основе. Данные о внешних факторах, оказывающих влияние на дебит нефти, доступны для прошлых периодов, но не доступны для прогнозного периода.

Показатели работы скважин, получаемые с датчиков или вносимые в информационные системы людьми, могут содержать ошибки и пропуски, поэтому важное значение приобретает обработка данных и фильтрация выбросов (так называемые point outliers).

При возникновении неполадок характер работы скважины значительно и быстро меняется, что осложняет прогнозирование дебита на отдаленные временные периоды.

**1. Формат ввода**

Вам предоставляются следующие файлы:

* "train.csv" – файл с данными для обучения моделей.

Целевая переменная: "Дебит нефти"

Характеристики тренировочной выборки:

* Размер train-таблицы: 67136 строк на 20 столбцов
* Количество уникальных скважин в train-таблице: 106

| Столбец | Описание |
| --- | --- |
| datetime | Дата |
| Номер скважины | Уникальный номер-идентификатор скважины |
| Дебит нефти | Прогнозируемая переменная – мера поступающей из скважины нефти за период |
| Давление забойное | Давление в стволе скважины с поправкой на глубину приведения забойного давления |
| x | Координата расположения скважины |
| y | Координата расположения скважины |
| Объем жидкости | Объем флюида (нефти и воды), поступающей из скважины нефти за период |
| Объем нефти | Объем нефти, поступающей из скважины нефти за период |
| Активная мощность (ТМ) | Активная мощность скважины по данным телеметрии |
| Время работы (ТМ) | Время работы по данным телеметрии |
| Газовый фактор рабочий (ТМ) | Количество нефтяного газа, приведенное к стандартным условиям, отнесенное к одной тонне добытой нефти |
| Давление буферное | Динамическое давление в скважине у устья во время ее фонтанирования нефтью, газом или водой |
| Давление забойное от Hд | Забойное давление от динамического уровня в скважине |
| Давление забойное от Pпр | Забойное давление от давления начала приемистости |
| Давление линейное (ТМ) | Линейное давление нефтяной скважины по данным телеметрии |
| Давление на входе ЭЦН (ТМ) | Давление на входе в погружной насос |
| Дебит газа (ТМ) | Дебит газа по данным телеметрии |
| Дебит газа попутного | Дебит попутного газа |
| Дебит жидкости (ТМ) | Дебит жидкости по данным телеметрии |
| Коэффициент мощности (ТМ) | Коэффициент мощности по данным телеметрии |

Для каждой скважины в train-файле представлены временные ряды длиной не менее 620 элементов. Некоторые скважины содержат большее количество элементов. Дискретность временных рядов - суточная.

Столбец с целевой переменной не содержит пропусков в данных, однако некоторые интервалы могли быть линейно интерполированы, что видно при построении графиков: интерполированные сегменты временных рядов выглядят как нехарактерно ровная линия на графике.

Столбцы с другими параметрами скважин могут содержать пропуски. При построении модели вы можете использовать все имеющиеся данные, или только часть параметров, или только предшествующие значения целевой переменной – на ваше усмотрение.

**1.2. Формат вывода**

Ответ принимается в формате .csv. Пример формируемого файла - "baseline\_forecast.csv".

Ваша модель должна сформировать прогнозы для всех 106 скважин на 90 дней вперед. csv файл должен содержать столбцы "datetime", "forecast" и "Номер скважины". Предсказания модели должны быть сформированы для периода c "1992-04-11" по "1992-07-09" – всего 90 дней. Пример генерации файла с ответом в базовом решении в файле "baseline\_solution.py".

|  |  |
| --- | --- |
| Столбец | Описание |
| datetime | Дата прогноза |
| forecast | Прогнозируемый дебит нефти |
| Номер скважины | Уникальный номер-идентификатор скважины |

**1.3. Описание обозначений**

ТМ – телеметрия

Нд – динамический уровень в скважине

Рпр – давление начала приемистости

ЭЦН – наиболее широко распространенный в России аппарат механизированной добычи нефти (ESP – electric submersible pump – погружной насос)

**1.4. Комментарии и оценка решений**

Решение задачи должно давать на выходе файл, содержащий прогноз для всех 106 скважин в train-выборке на 90 дней вперед.

Ошибка прогноза будет оцениваться на отложенной выборке, равной одному горизонту прогнозирования (90 дней). Для оценки и ранжирования предлагаемых участниками моделей будет использоваться показатель RMSE (Root Mean Square Error: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean\_squared\_error.html)

Финальная метрика получается усреднением метрик, полученных по всем скважинам.

**1.5. Идея решения**

Базовое решение, основанное на оптимизации параметров, описывающих кривую падения дебита нефти, представлено в файле "baseline\_solution.py". В этом решении для построения прогноза используются только предшествующие значения целевой переменной.

Подробнее о DCA (Decline Curve Analysis): https://petrowiki.spe.org/Production\_forecasting\_decline\_curve\_analysis.

Для решения задачи возможно использование любых моделей и подходов к прогнозированию временных рядов, в том числе алгоритмов AutoML, классических моделей МО, нейронных сетей. Участники могут использовать различные методы заполнения пропусков в данных либо не учитывать параметры, содержащие пропуски. Допускается построение моделей на основе любых наборов параметров, доступных в файле "train.csv".

Доступные в train-выборке данные не позволяют напрямую определить этап жизненного цикла скважины, однако на основе координат расположения скважин и укрупненных показателей их работы можно объединить схожие скважины в кластеры.

Размер скользящего окна для обучения модели и получения прогнозов может быть выбран участниками самостоятельно и включать от нескольких дней, непосредственно предшествующих прогнозному периоду, до 620 дней (гарантированная протяженность временных рядов, доступная по всем скважинам в выборке).

На практике характерные паттерны изменений ключевых показателей при дефектах скважин проявляются на горизонте в 15-20 дней, для оценки общего тренда дебита нефти необходимо использовать более длительные интервалы доступных данных.