

# ПОДГОТОВКА ДАННЫХ СКВАЖИНЫ НЕФТЕГАЗОВОГО МЕСТОРОЖДЕНИЯ ДЛЯ ОБУЧЕНИЯ МОДЕЛИ НЕЙРОННОЙ СЕТИ С ЦЕЛЬЮ ПРОГНОЗА ДЕБИТА НЕФТИ

*Е.И. Губин, к.т.н., доц. ОИТ ИШИТР,  
С.Н. Ушаков, студент гр. 8ПМ9И  
М.А. Сальников, студент гр. 8ПМ9И  
Т.А. Михеев, студент гр. 2БМ95  
Томский политехнический университет  
E-mail: wowman9721@gmail.com*

## Введение

В настоящее время является актуальной проблема разработки виртуального расходомера, позволяющего оценивать и прогнозировать дебит (расход) нефти по набору признаков. В настоящее время разработано несколько методов для оценки нефтяного дебита, которые различаются по своей методологической природе и характеристикам [1].

Обычно используются числовые модели для расчета дебита с использованием доступных полевых измерений, таких как давление и температура. Однако для более эффективной оценки дебита необходимо делать анализ на основе большего количества признаков [2].

Целью работы является подготовка данных для обучения модели с целью прогноза дебита нефти на основе замеров нескольких параметров.

## Описание процесса подготовки данных

В качестве исходных данных имеется 2 датасета:

- данные замеров АГЗУ (автоматическая групповая замерная установка), которые снимаются раз в 1-1,5 суток. Фрагмент замеров АГЗУ представлен на рисунке 1;
- данные замеров СУ (система управления), которые снимаются каждые 7-12 секунд. Фрагмент замеров СУ представлен на рисунке 2.

	MEASDATE	ID	WELL_ID	MEASDATE.1	ACURR	ABVOL	POWERFACTOR	LOADFACTOR	ESPWINDINGINSRESISTANCE	OUTFREQ
0	2020-06-22 15:22:28	175973454	8020262100	2020-06-22 15:22:28	NaN	NaN	NaN	60.0	NaN	NaN
1	2020-06-22 15:22:29	175973455	8020262100	2020-06-22 15:22:29	NaN	NaN	NaN	60.0	NaN	NaN
2	2020-06-22 15:22:40	175973456	8020262100	2020-06-22 15:22:40	NaN	NaN	NaN	60.0	NaN	NaN
3	2020-06-22 15:22:41	175973457	8020262100	2020-06-22 15:22:41	NaN	NaN	NaN	60.0	NaN	NaN
4	2020-06-22 15:22:52	175973458	8020262100	2020-06-22 15:22:52	NaN	NaN	NaN	60.0	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
614717	2021-01-18 01:33:53	645466652	8020262100	2021-01-18 01:33:53	21.5	383.0	0.63	60.0	9999.0	60.0
614718	2021-01-18 01:34:06	645466653	8020262100	2021-01-18 01:34:06	21.0	383.0	0.63	59.0	9999.0	60.0
614719	2021-01-18 01:34:19	645466654	8020262100	2021-01-18 01:34:19	21.0	383.0	0.63	59.0	9999.0	60.0
614720	2021-01-18 01:34:31	645466655	8020262100	2021-01-18 01:34:31	21.0	383.0	0.63	60.0	9999.0	60.0
614721	2021-01-18 01:34:32	645466656	8020262100	2021-01-18 01:34:32	21.0	383.0	0.63	60.0	9999.0	60.0

Рис. 1. Замеры СУ

	ID	WELL_ID	MEASDATE	PLANDURATION	FACTDURATION	FLOWRATE	GASFLOWRATE	GZUMEASPRESSURE
0	178727071	8020262100	2020-06-24 07:18:32	360	360	35	251.0	13.785
1	181164696	8020262100	2020-06-25 18:31:36	360	360	34	316.0	13.625
2	182242207	8020262100	2020-06-27 05:44:44	360	360	34	332.0	13.370
3	182242468	8020262100	2020-06-28 16:57:47	360	360	35	323.0	13.620
4	182242747	8020262100	2020-06-30 04:10:51	360	360	35	324.0	13.795
...	...	...	...	...	...	...	...	...
170	623311016	8020262100	2021-01-10 19:48:29	240	240	39	289.0	13.705
171	627444619	8020262100	2021-01-12 03:00:38	240	240	39	315.0	13.675
172	631341265	8020262100	2021-01-13 10:12:36	240	240	40	339.0	13.075
173	637205844	8020262100	2021-01-15 08:41:17	240	125	39	266.0	13.385
174	640414443	8020262100	2021-01-16 09:45:49	240	240	39	273.0	13.820

Рис. 2. Замеры АГЗУ

Для начала необходимо сформировать датафрейм с ключевыми признаками на основе датасета с замерами АГЗУ. Таковыми признаками являются:

- POWERFACTOR - коэффициент мощности насоса;
- WORKINGFREQ - рабочая частота насоса;
- LOADFACTOR - коэффициент загрузки насоса;
- OUTFREQ - выходная частота насоса;
- PUMPINTAKEPRESSURE - давление на приеме насоса;
- ANNULUSPRESSURE - затрубное давление;
- MEASDATE - дата окончания замера.

Для обучения модели все данные должны быть загружены в нейронную сеть в числовом виде. Однако значения параметра MEASDATE представлены в виде дат, поэтому их необходимо преобразовать в численный тип с помощью метода временной метки (timestamp). Далее необходимо разбить дату (MEASDATE) из замеров АГЗУ (рис.2) на интервалы по 10 минут (600 секунд), и, согласно этим интервалам, взять ближайшие значения замеров СУ. Подготовка данных ведется для возможности использования модели машинного обучения с учителем. В качестве учителя будет выступать суточный дебит скважины (FLOWRATE), который необходимо добавить в получившийся датафрейм в качестве отдельной колонки. Фрагмент подготовленных данных представлен на рисунке 3.

	1	2	3	4	5	7	8 FLOWRATE
0	[61.6161616161616161, 61.6161616161616161, 62.6262...	[95.750000000000001, 95.750000000000001, 96.6666...	[73.17073170731707, 73.17073170731707, 74.3902...	[95.750000000000001, 95.750000000000001, 96.6666...	[1.2190661952944026, 0.6217237596001439, 0.377...	[91.47339075897742, 91.88266767512498, 91.6780...	[1592896716.0, 1592897308.0, 1592897916.0, 159...
1	[62.62626262626262, 62.62626262626262, 0.0, 0,...	[96.66666666666667, 96.66666666666667, 0.0, 95...	[74.39024390243902, 73.17073170731707, 0.0, 73...	[96.66666666666667, 96.66666666666667, 0.0, 95...	[1.462879434353276, 1.0971595757649624, 2.1821...	[91.74624203640913, 91.74624203640913, 92.3601...	[1593023493.0, 1593024101.0, 1593024698.0, 159...
2	[62.62626262626262, 62.62626262626262, 62.6262...	[96.66666666666667, 95.91666666666667, 95...	[0.0, 2.4390243902439024, 69.51219512195122, 7...	[0.0, 9.616666666666667, 95.91666666666667, 95...	[0.0, 2.0602218700475348, 3.62062660002438, 3.14519...	[91.47339075897742, 91.20053948154572, 90.3819...	[1593150281.0, 1593150879.0, 1593151487.0, 159...
3	[0.0, 0.0, 0.0, 61.6161616161616161, 62.62626262...	[94.81666666666668, 95.78333333333333, 95.7833...	[69.51219512195122, 70.73170731707317, 71.9512...	[94.81666666666668, 95.78333333333333, 95.7833...	[3.9863464586126938, 3.7425332195538132, 3.510...	[89.35879335888175, 89.29058053952383, 89.2905...	[1593277065.0, 1593277663.0, 1593278263.0, 159...

Рис. 3. Фрагмент подготовленных данных

### Заключение

В результате подготовки данных был сформирован датафрейм, строки которого представляют собой массивы данных одинаковой длины. Одинаковые размеры массивов важно учитывать при корректной загрузке данных в нейронную сеть.

### Список использованных источников

1. Aarsnes, U., Flatten, T., Aamo, O., 2016. Review of two-phase flow models for control and estimation. Annu. Rev. Contr. 42, 50–62.
2. Andrianov, N., 2018. A machine learning approach for virtual flow metering and forecasting. IFAC-Pap. Online 51, 191–196.