Содержание

# Запрос в SQL

SELECT df.ID, TIMESTAMP, dj.DESCRIPTION, VALUE FROM DRILL\_GEO\_GAUGE\_TRANS df   
LEFT JOIN DRILL\_GEO\_GAUGE\_TYPE dj ON dj.ID = df.GAUGE\_TYPE\_ID   
ORDER BY df.ID;

# Характеристика компьютера

OS Name: Майкрософт Windows 11 Домашняя

OS Version: 10.0.22000 22000

CPU: Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz

RAM: 7.891834259033203 GB

Graphics Card: Intel(R) UHD Graphics 630

# Первичный осмотр данных

Всего имеется 3253056 наблюдений.  
Имееются следующие переменные:

1. Счетчик штанг
2. Частота вращения шпинделя
3. Индекс углубки
4. Мгновенная механическая скорость бурения
5. Уставка расхода промывочного насоса (производительность)
6. Глубина скважины (м)
7. Глубина обсадки
8. Давление масла двигателя
9. Частота вращения коленвала двигателя
10. Напряжение бортовой сети
11. Расход топлива мгновенный в час по шине CAN
12. Суммарный расход топлива (накопительный)
13. Крутящий момент на шпинделе
14. Моточасы работы двигателя
15. Температура редуктора вращателя
16. Усилие в полости подпора
17. Давление промывочной жидкости
18. Давление в магистрали главного гидронасоса
19. Усилие в полости подачи гидроцилиндра
20. Температура охлаждающей жидкости
21. Расход промывочной жидкости
22. Температура гидравлического масла
23. Рабочий замер
24. Осевая нагрузка
25. Вес снаряда
26. Мертвый замер
27. Глубина по замеру лебедки ССК
28. Сменная проходка

|  |  |
| --- | --- |
| Название переменной | Количество пропусков в изначальных данных |
| # | 0 |
| ID | 0 |
| TIMESTAMP | 0 |
| DESCRIPTION | 0 |
| VALUE | 0 |

Количество дубликатов в изначальных данных: 0 строчки

|  |  |
| --- | --- |
| Название переменной | Количество пропусков в данных после обработки (приведенные к широкому формату) |
| TIMESTAMP | 0 |
| Глубина скважины (м) | 0 |
| Давление промывочной жидкости | 0 |
| Индекс углубки | 0 |
| Крутящий момент на шпинделе | 0 |
| Осевая нагрузка | 0 |
| Расход промывочной жидкости | 0 |
| Расход топлива мгновенный в час по шине CAN | 0 |
| Частота вращения шпинделя | 0 |

Количество дубликатов в данных после обработки: 2608 строчки

## Детекция выбросов

В таблице предоставленны данные о выбросах

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Название метода\Название переменной | Глубина скважины (м) | Давление промывочной жидкости | Индекс углубки | Крутящий момент на шпинделе | Осевая нагрузка | Расход промывочной жидкости | Расход топлива мгновенный в час по шине CAN | Частота вращения шпинделя |
| z-score (абсолютный) | 4228 | 2904 | 2883 | 403 | 3719 | 1672 | 104 | 145 |
| z-score (в процентах) | 4.8 % | 3.3 % | 3.3 % | 0.5 % | 4.2 % | 1.9 % | 0.1 % | 0.2 % |
| IQR (абсолютный) | 18737 | 1643 | 12128 | 5765 | 2795 | 1035 | 93 | 297 |
| IQR (в процентах) | 21.2 % | 1.9 % | 13.7 % | 6.5 % | 3.2 % | 1.2 % | 0.1 % | 0.3 % |

# Результаты машинного обучения

## Линейные модели

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| estimator | mse | rmse | r2 | mae | best\_estimator | time |
| LinearRegression | 2075332.7 | 1440.6 | 0.04 | 317.2 | - | 0.03418087959289551 |
| Lasso | 2075319.1 | 1440.6 | 0.04 | 317.2 | - | 0.0323328971862793 |
| Lasso opt | 2075319.1 | 1440.6 | 0.04 | 317.2 | Lasso(alpha=1) | 1.1559443473815918 |
| Ridge | 2075332.7 | 1440.6 | 0.04 | 317.2 | - | 0.029036760330200195 |
| Ridge opt | 2075332.2 | 1440.6 | 0.04 | 317.2 | Ridge(alpha=91) | 0.8146779537200928 |

## Деревянные модели

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| estimator | mse | rmse | r2 | mae | best\_estimator | time |
| DecisionTreeRegressor | 1761677.3 | 1327.3 | 0.18 | 57.6 | - | 0.9011285305023193 |
| DecisionTreeRegressor opt | 437391.7 | 661.4 | 0.8 | 176.0 | DecisionTreeRegressor(max\_depth=3) | 1.2738885879516602 |
| RandomForestRegressor | 76633.4 | 276.8 | 0.96 | 34.1 | - | 52.6312575340271 |
| RandomForestRegressor opt | 122329.4 | 349.8 | 0.94 | 111.4 | RandomForestRegressor(max\_depth=7, n\_estimators=300) | 123.4983229637146 |

## Модели бустинга

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| estimator | mse | rmse | r2 | mae | best\_estimator | time |
| GradientBoostingRegressor | 714232.6 | 845.1 | 0.67 | 88.2 | - | 10.896440744400024 |
| GradientBoostingRegressor opt | 38422.0 | 196.0 | 0.98 | 60.5 | GradientBoostingRegressor(n\_estimators=500) | 257.5199055671692 |
| <catboost.core.CatBoostRegressor object at 0x00000216C57A2190> | 1155815.7 | 1075.1 | 0.46 | 59.3 | - | 13.341974258422852 |