



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Исследование датасета «Оставшееся количество циклов зарядки аккумулятора» (ориг. *Battery Remaining Useful Life - RUL*)

Выпускная квалификационная работа
1 мая 2023 г.

Ларин П. М.



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Информация о датасете RUL

Автор датасета: **Ignacio Viñuales**

Оригинальное название: **Battery Remaining Useful Life (RUL)**

Домашняя страница датасета: https://github.com/ignavinuales/Battery_RUL_Prediction

На Каггле: <https://www.kaggle.com/datasets/ignaciovinuales/battery-remaining-useful-life-rul>

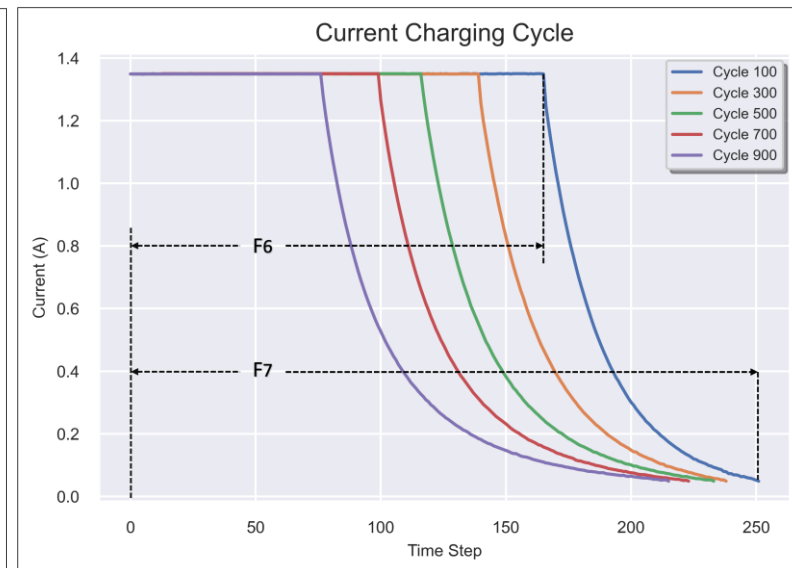
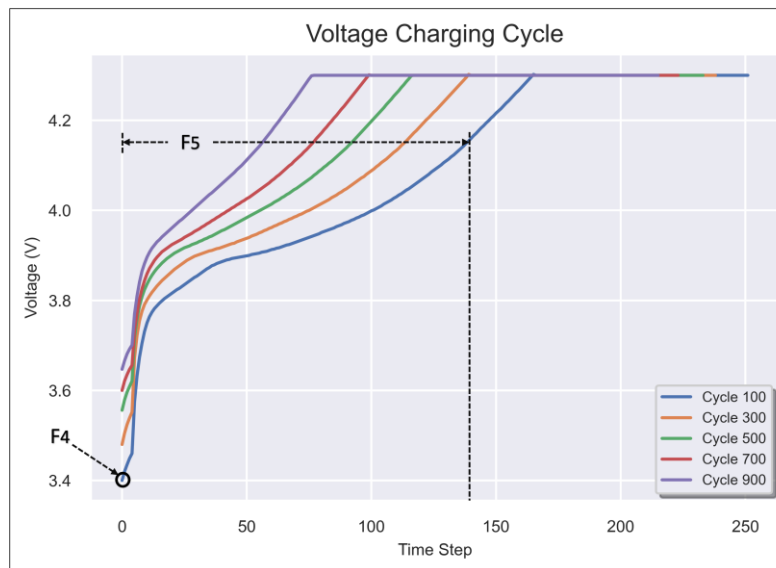
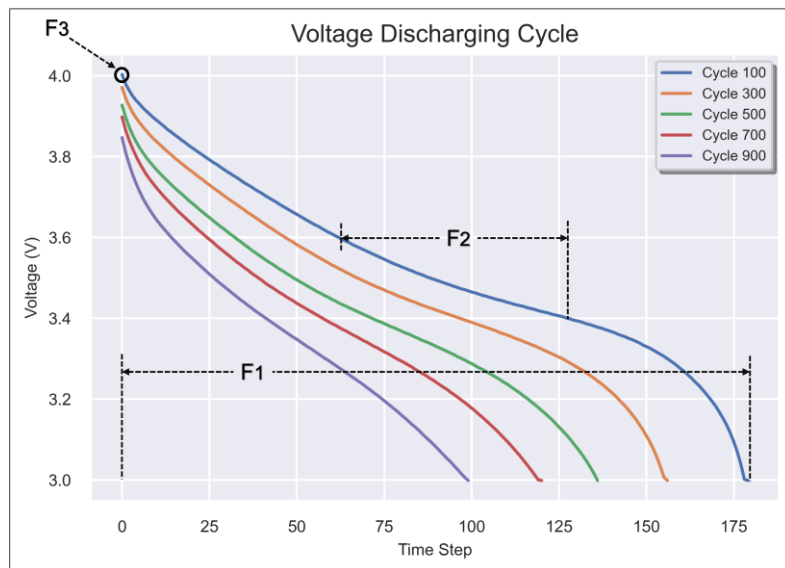
Описание автора датасета

Гавайский Институт природной энергии протестировал 14 аккумуляторов типа NMC-LCO 18650 номинальной емкости 2.8 А·ч. Каждый аккумулятор был подвергнут более чем 1000 циклам разрядки/зарядки при температуре 25°C, включающим CC-CV зарядку в режиме C/2 и разрядку в режиме 1.5C.

На основе соответствующего датасета я [автор] создал новые признаки, описывающие динамику напряжения и силы тока для каждого цикла. Эти признаки можно использовать для предсказания оставшегося количества циклов аккумулятора. Датасет содержит сводную информацию о 14 аккумуляторах.



Информация о датасете RUL



Признаки

- Индекс цикла
- F1: Полная продолжительность разрядки (с)
- F2: Продолжительность разрядки с 3.6 В до 3.4 В (с)
- F3: Стартовое напряжение при разрядке (В)
- F4: Стартовое напряжение при зарядке (В)
- F5: Продолжительность зарядки до 4.15 В (с)
- F6: Продолжительность зарядки постоянным током (с)
- F7: Полная продолжительность зарядки (с)
- RUL: Целевая переменная (оставшееся количество циклов зарядки аккумулятора)



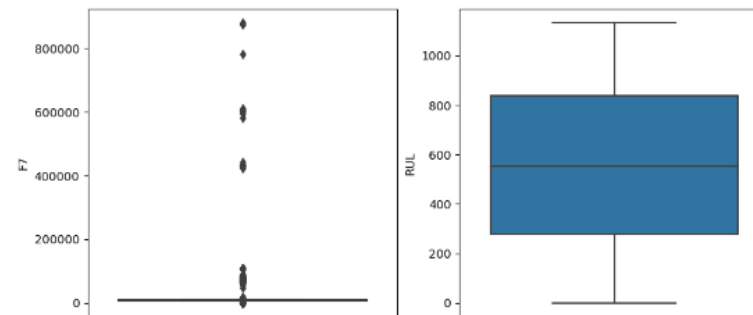
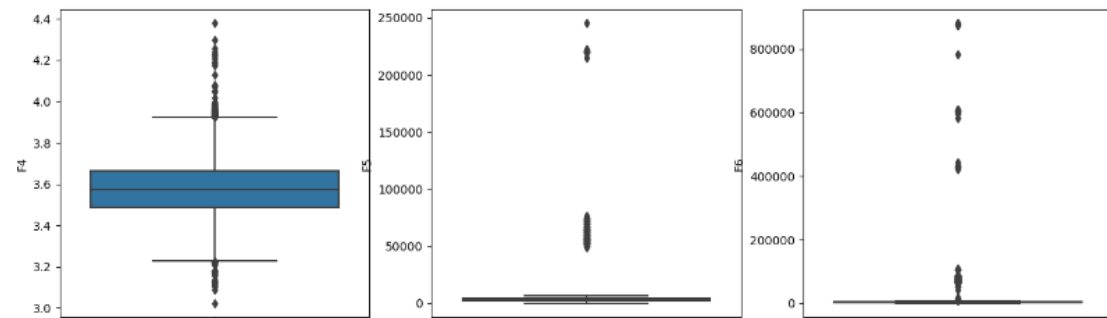
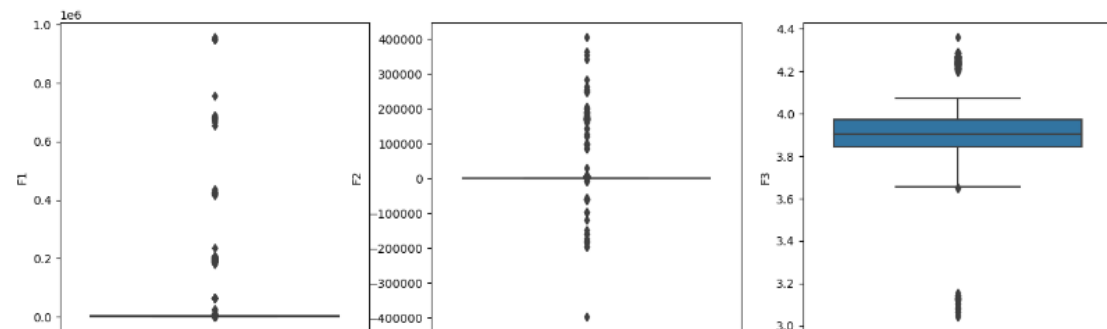
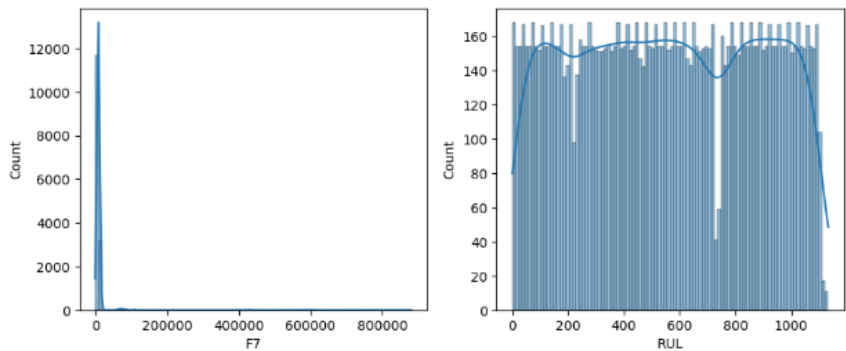
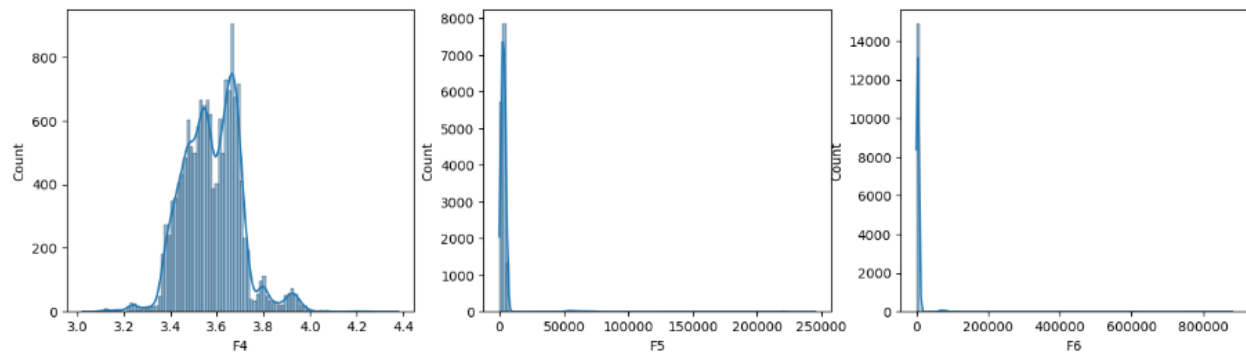
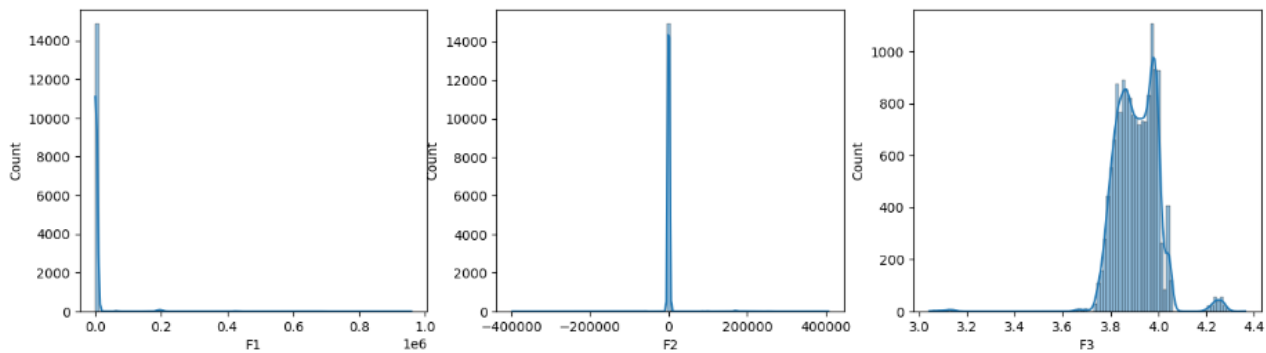
Информация о датасете RUL

Размер исходного
датасета: (15064, 9)

Cycle_Index	Discharge Time (s)	Decrement 3.6-3.4V (s)	Max. Voltage Dischar. (V)	Min. Voltage Charg. (V)	Time at 4.15V (s)	Time constant current (s)	Charging time (s)	RUL
1	2595.3	1151.4885	3.67	3.211	5460.001	6755.01	10777.82	1112
2	7408.64	1172.5125	4.246	3.22	5508.992	6762.02	10500.35	1111
3	7393.76	1112.992	4.249	3.224	5508.993	6762.02	10420.38	1110
4	7385.5	1080.320667	4.25	3.225	5502.016	6762.02	10322.81	1109
6	65022.75	29813.487	4.29	3.398	5480.992	53213.54	56699.65	1107
7	3301.18	1194.235077	3.674	3.504	5023.633636	5977.38	5977.38	1106
8	5955.3	1220.135329	4.013	3.501	5017.495	5967.55	5967.55	1105
9	5951.2	1220.135329	4.014	3.501	5017.496	5962.21	5962.21	1104
10	5945.44	1216.920914	4.014	3.501	5009.993667	5954.91	5954.91	1103
11	435251.49	263086.078	4.267	3.086	269.984	443700.02	443700.02	1102
12	3228.58	1135.349333	3.689	3.485	5033.075692	5969.89	5969.89	1101
13	6019.9	1058.279724	4.045	3.475	5053.842846	5980.77	5980.77	1100
14	6026.59	1049.487845	4.047	3.477	5046.4295	5966.82	5966.82	1099
15	6008.07	1065.372059	4.045	3.48	5033.075769	5954.47	5954.47	1098
16	423271.35	168773.265	4.27	3.108	219923.996	430028.84	430028.84	1097
17	2261.34	883.2	4.038	3.901	1949.664	2922.69	6070.11	1096
18	2259.46	883.199	4.042	3.373	5181.377	6161.38	9310.98	1095
19	2256.61	878.4	4.042	3.374	5181.375	6154.37	9296.64	1094
20	2252.83	873.601	4.043	3.374	5174.334	6147.33	9243.58	1093



Разведочный анализ





Выбросы

	F1	F2	F3	F4	F5	F6	F7	RUL
10	3228.58	1135.349333	3.689	3.485	5033.075692	5969.89	5969.89	1101
11	6019.90	1058.279724	4.045	3.475	5053.842846	5980.77	5980.77	1100
12	6026.59	1049.487845	4.047	3.477	5046.429500	5966.82	5966.82	1099
13	6008.07	1065.372059	4.045	3.480	5033.075769	5954.47	5954.47	1098
14	423271.35	168773.265000	4.270	3.108	219923.996000	430028.84	430028.84	1097
15	2261.34	883.200000	4.038	3.901	1949.664000	2922.69	6070.11	1096
16	2259.46	883.199000	4.042	3.373	5181.377000	6161.38	9310.98	1095
17	2256.61	878.400000	4.042	3.374	5181.375000	6154.37	9296.64	1094
18	2252.83	873.601000	4.043	3.374	5174.334000	6147.33	9243.58	1093

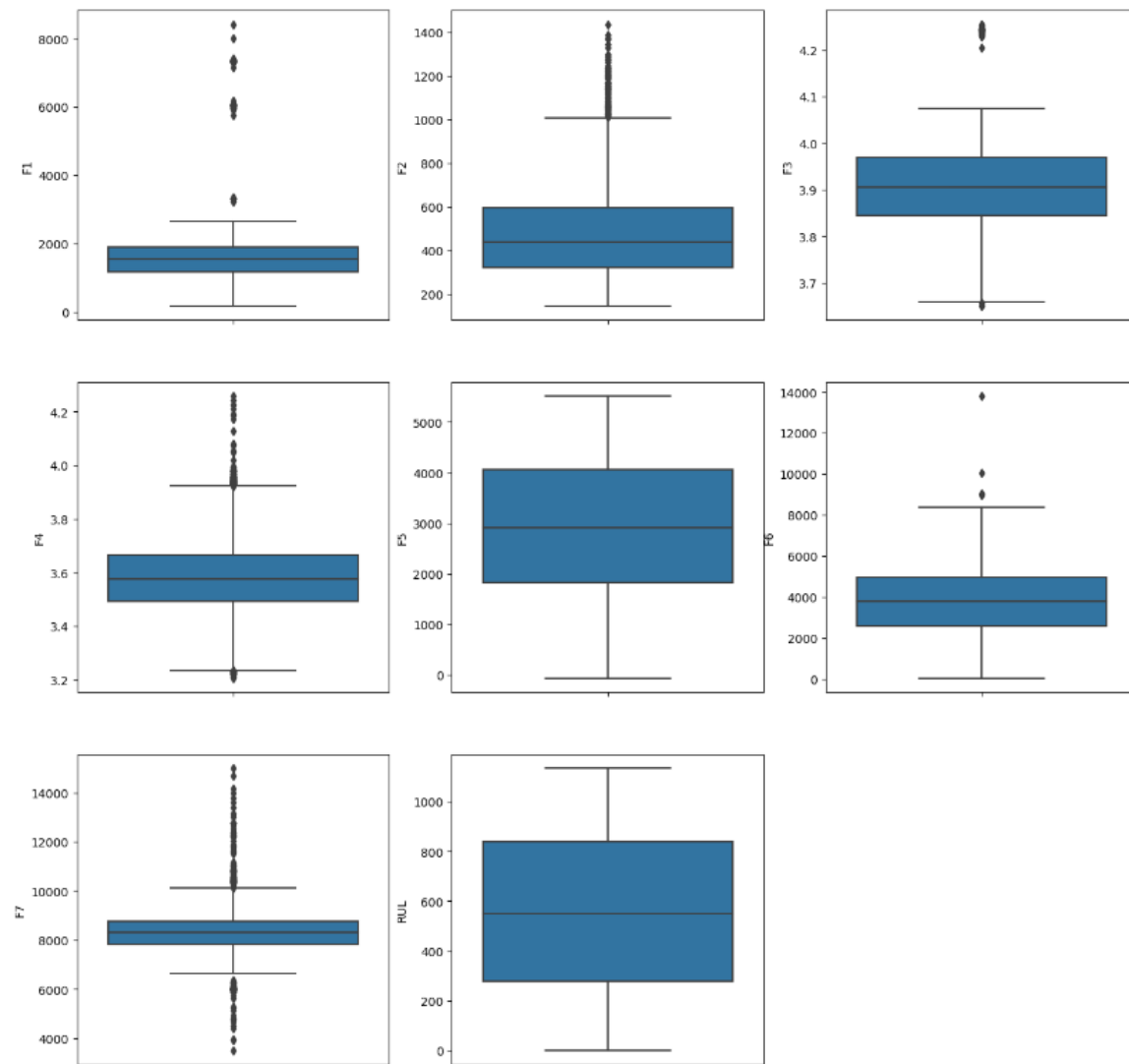
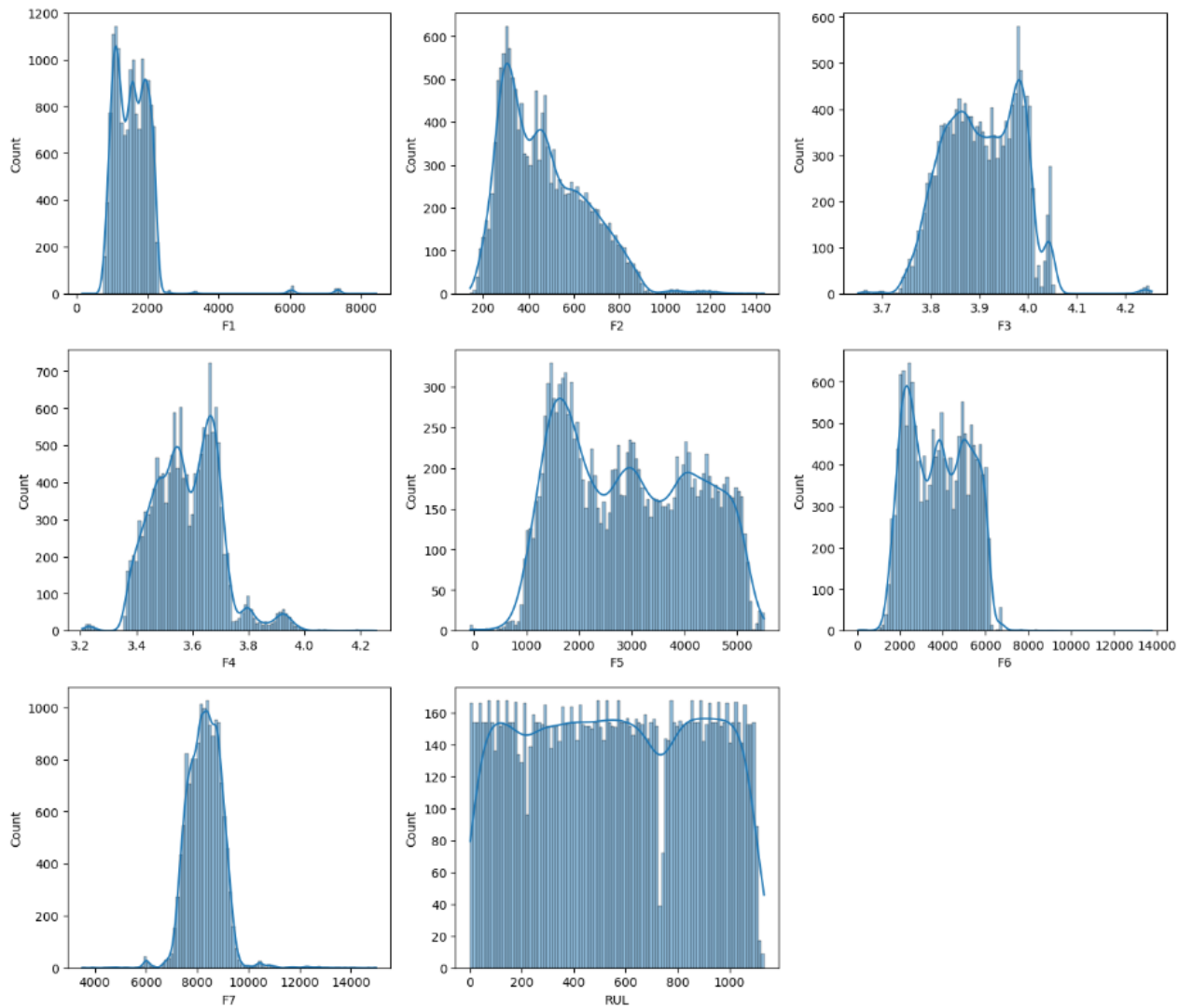
	F1	F2	F3	F4	F5	F6	F7	RUL
11822	1017.06	273.371429	3.800	3.674	1457.375000	2096.38	7823.50	41
11823	1014.88	273.500000	3.800	3.673	1452.875000	2096.38	7793.94	40
11824	1008.00	272.457143	3.800	3.675	1448.375000	2060.38	7795.12	39
11825	1012.25	271.542857	3.799	3.676	1443.875000	2060.38	7792.44	38
11826	32.38	42.105263	3.114	3.674	114.250000	8.00	8.00	37
11827	1014.69	271.600000	3.799	3.671	1529.375000	2168.38	7963.75	36
11828	1014.62	271.657143	3.799	3.673	1448.375000	2060.38	7863.06	35
11829	1009.75	269.828571	3.799	3.673	1443.875000	2060.38	7856.69	34
11830	1007.00	268.914286	3.798	3.674	1432.884429	2060.31	7783.56	33

	F1	F2	F3	F4	F5	F6	F7	RUL
7603	2188.81	818.375	4.044	3.384	5002.813	5965.38	9136.38	1049
7604	2186.38	818.438	4.044	3.385	4999.313	5958.31	9101.31	1048
7605	2186.38	818.375	4.045	3.385	4992.375	5958.38	9136.38	1047
7606	2184.00	806.375	4.026	3.705	4956.000	5922.00	9142.00	1046
7607	207013.99	-98271.841	4.272	3.159	73349.980	85080.00	85080.00	1045
7608	11755.36	807.000	4.005	3.380	5250.080	6200.48	9556.67	1044
7609	2184.00	814.400	4.008	3.389	5012.353	5948.35	9188.35	1043
7610	2183.74	816.000	4.009	3.388	5012.318	5948.32	9188.32	1042
7611	2172.00	816.000	4.009	3.389	5005.152	5948.35	9192.90	1041

Подобные выбросы невозможно интерпретировать как корректные результаты измерений, поэтому перед нами либо ошибки измерений, либо ошибки обработки. Данные выбросы необходимо удалить.

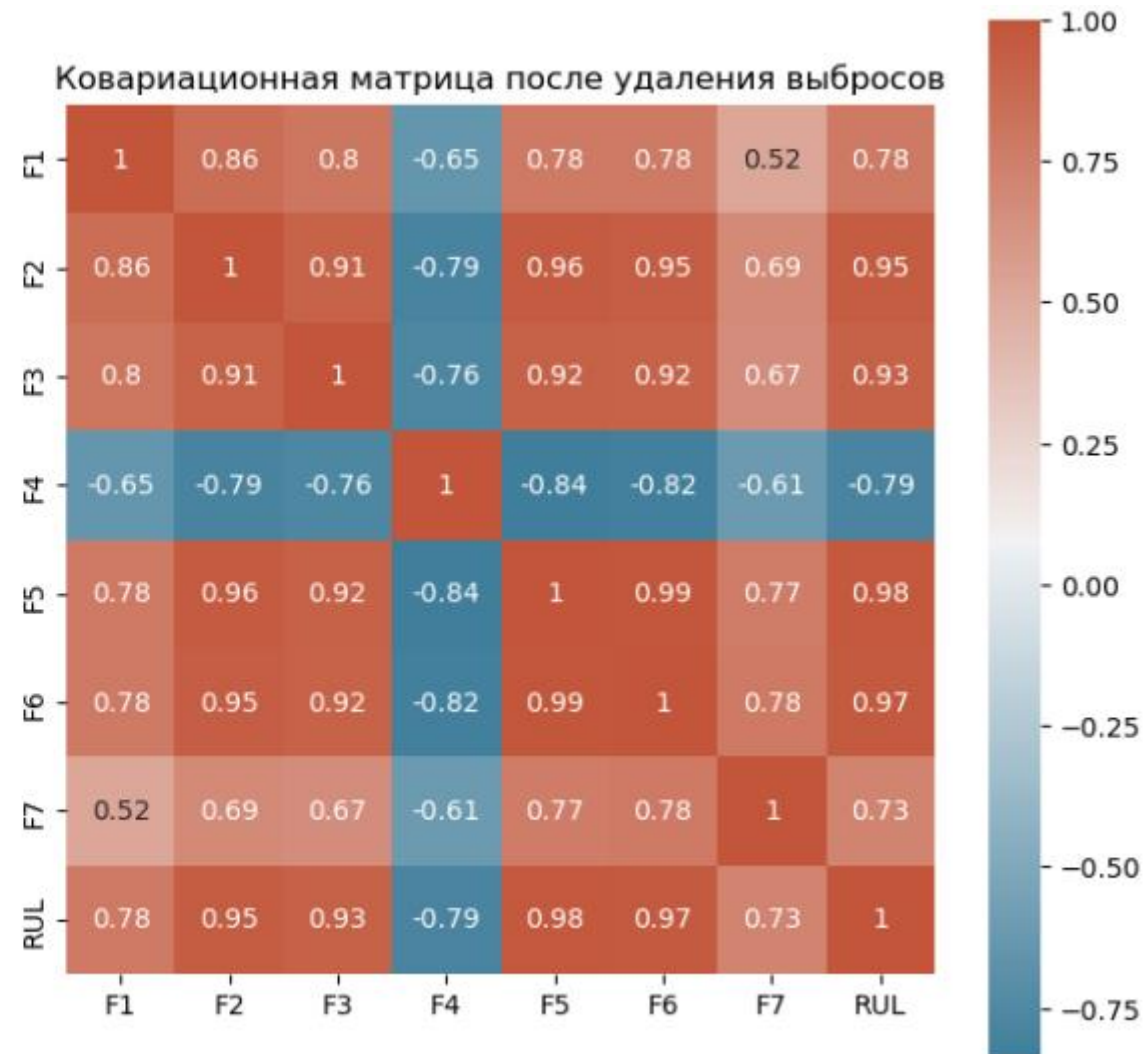
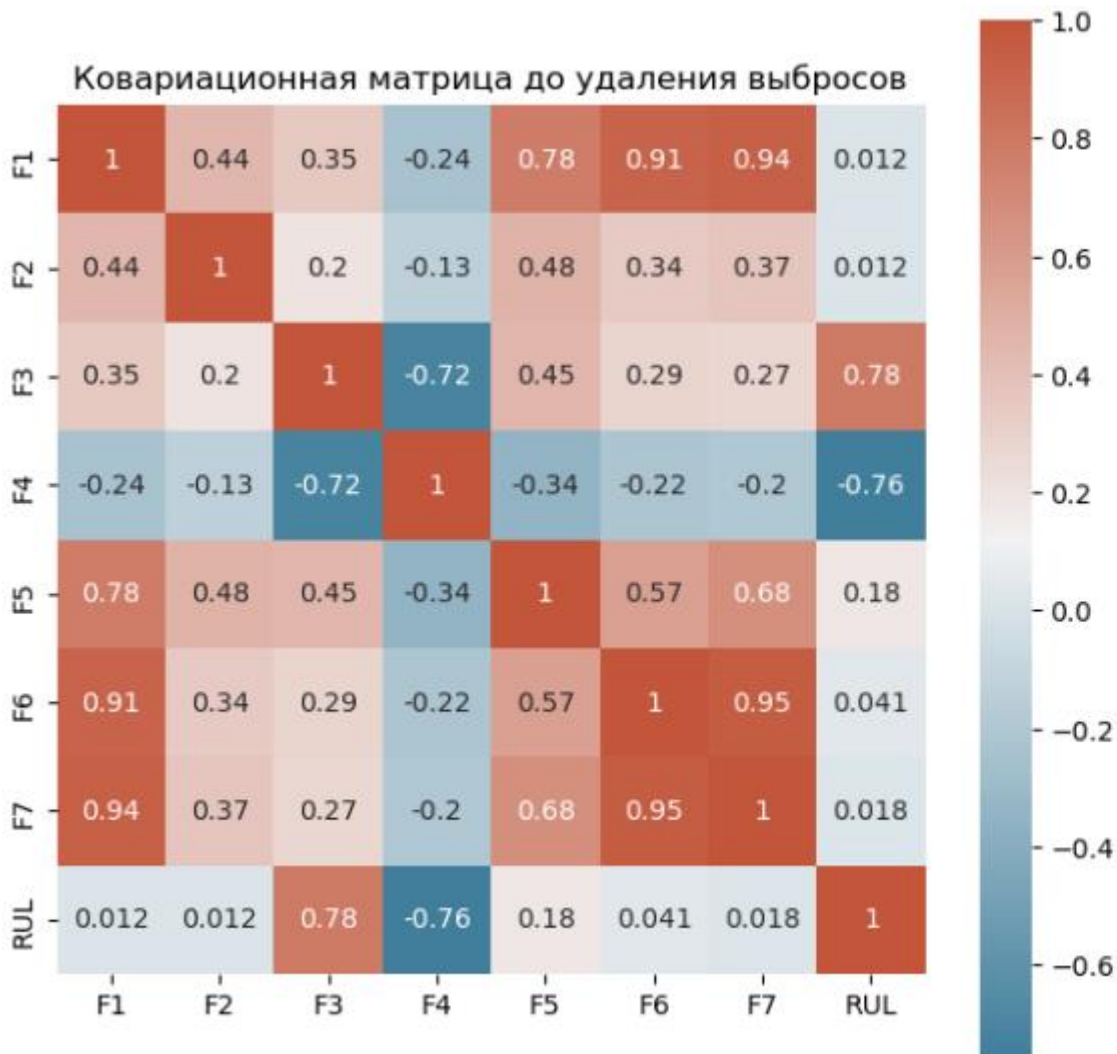


После удаления выбросов



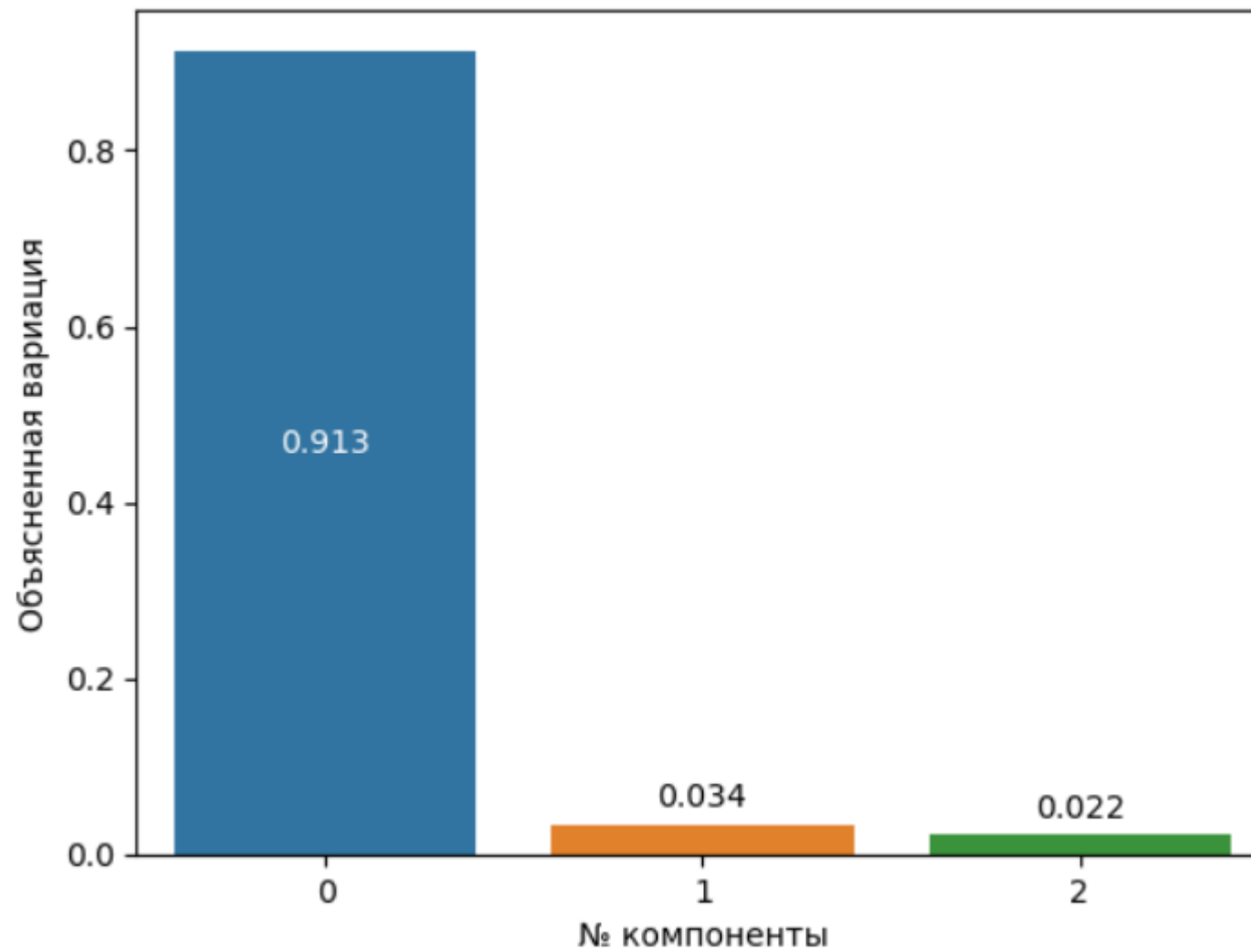


Ковариационные матрицы





Метод главных компонент





Регрессоры, метрики и методика

Используемые регрессоры

Scikit-learn

- Линейная регрессия
- Ridge (L_2 – регуляризация, или регуляризация Тихонова)
- Полиномиальная регрессия
- Метод k -ближайших соседей
- RandomForest
- Градиентный бустинг

TensorFlow

- Полносвязная нейронная сеть

Используемые метрики

Коэффициент детерминации (R^2)

- Лучший в тесте
- Лучший на 1 признаке
- Лучший на 2 признаках
- Лучший на 3 признаках

Средняя абсолютная ошибка (MAE)

- MAE, соответствующая каждому из вышеприведенных R^2

Методика

Для каждого регрессора и каждого $n \in [1, 7]$ проводится поиск n признаков, оптимальных по критерию R^2 из всех наборов из n признаков. На данных оптимальных наборах проводится кросс-валидация и по ее результатам – усреднение R^2 и MAE.

Из полученных таким образом наборов $\{R^2_n\}, \{MAE_n\}, n \in [1, 7]$, выбираются $\max \{R^2_n\}$ и $\min \{MAE_n\}$ в качестве лучших в тесте.

Для ансамблевых методов и kNN дополнительно проводится подбор гиперпараметров.



Пример проведения регрессии

Метод k-ближайших соседей

kNN с количеством соседей от 2 до 6:

```
In [40]: kNN1 = apply_regressor(KNeighborsRegressor(n_neighbors=2), df_X, dfc.RUL)
```

R2	RMSE	MAE	best features

0.9679	57.65	40.09	[1]
0.9729	52.92	35.23	[1 6]
0.9761	49.69	32.82	[1 5 6]
0.9862	37.75	20.74	[1 4 5 6]
0.9925	27.80	12.36	[1 2 4 5 6]
0.9950	22.66	8.03	[1 2 3 4 5 6]
0.9963	19.46	6.11	[ALL]

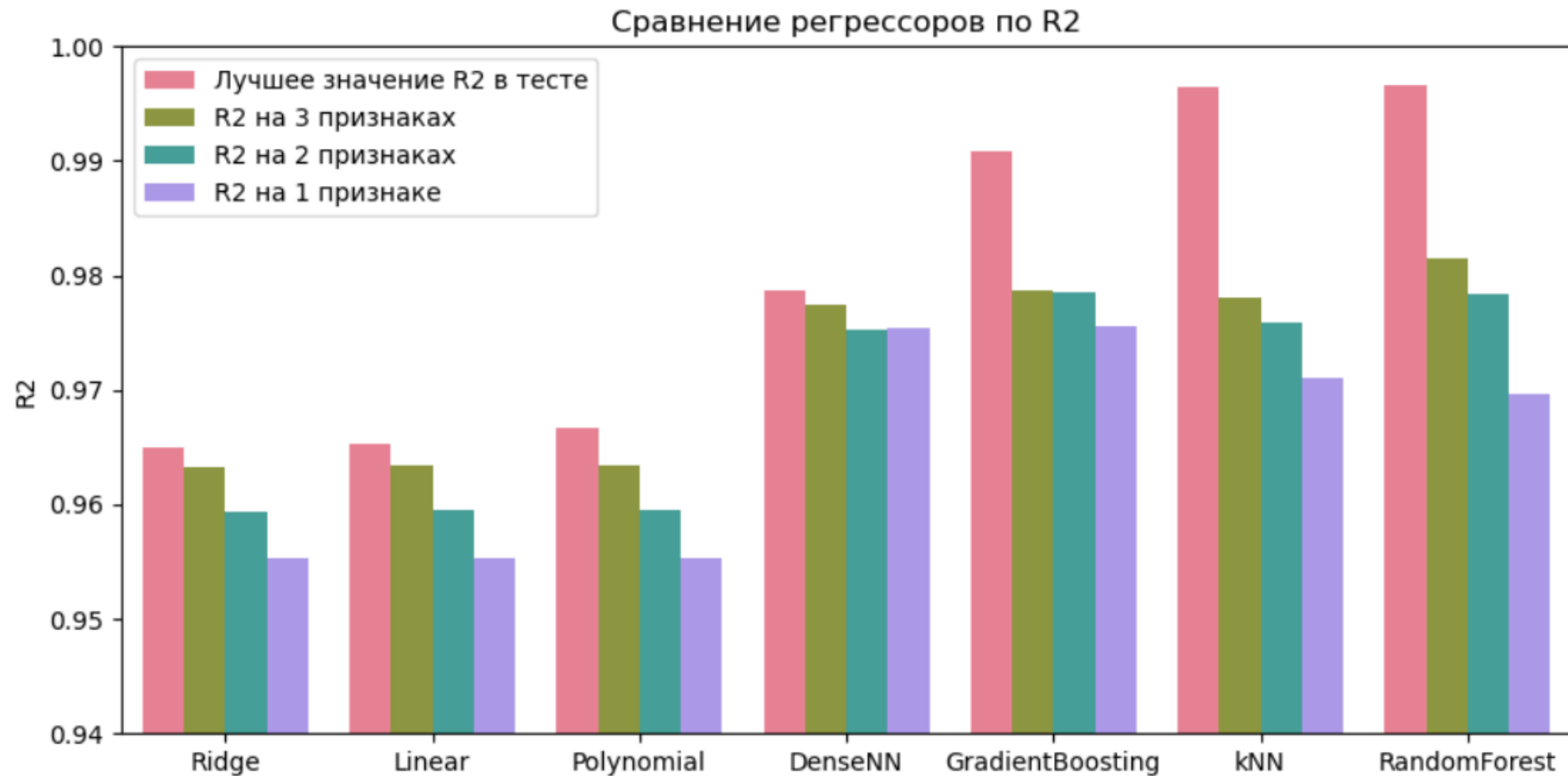
```
In [41]: kNN2 = apply_regressor(KNeighborsRegressor(n_neighbors=3), df_X, dfc.RUL)
```

R2	RMSE	MAE	best features

0.9711	54.65	38.35	[1]
0.9759	49.95	33.74	[1 6]
0.9781	47.60	32.26	[1 5 6]
0.9866	37.16	21.78	[1 4 5 6]
0.9926	27.64	13.35	[1 2 4 5 6]
0.9952	22.23	8.76	[1 2 3 4 5 6]
0.9965	19.10	6.59	[ALL]

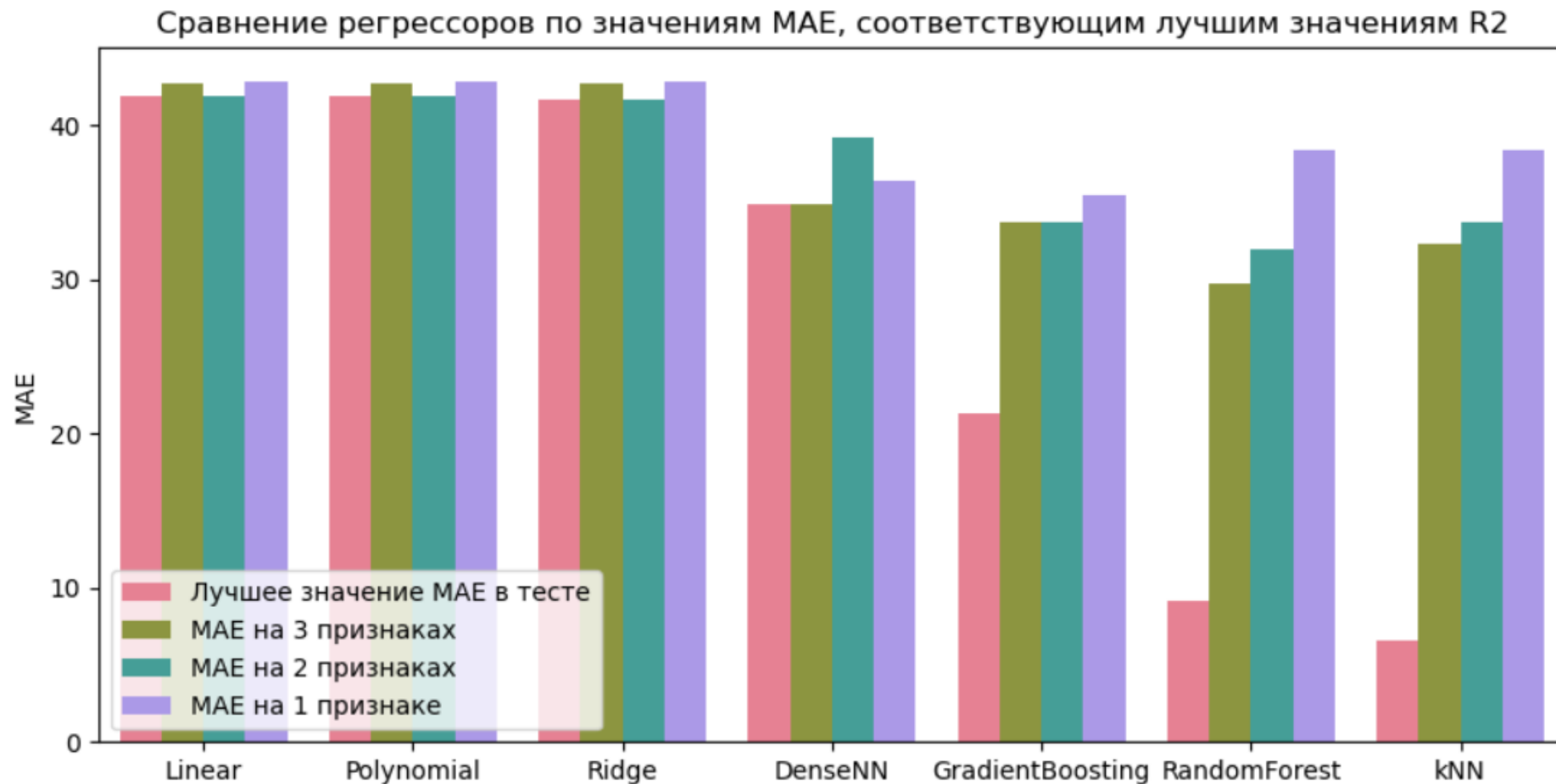


Результаты: R^2





Результаты: MAE





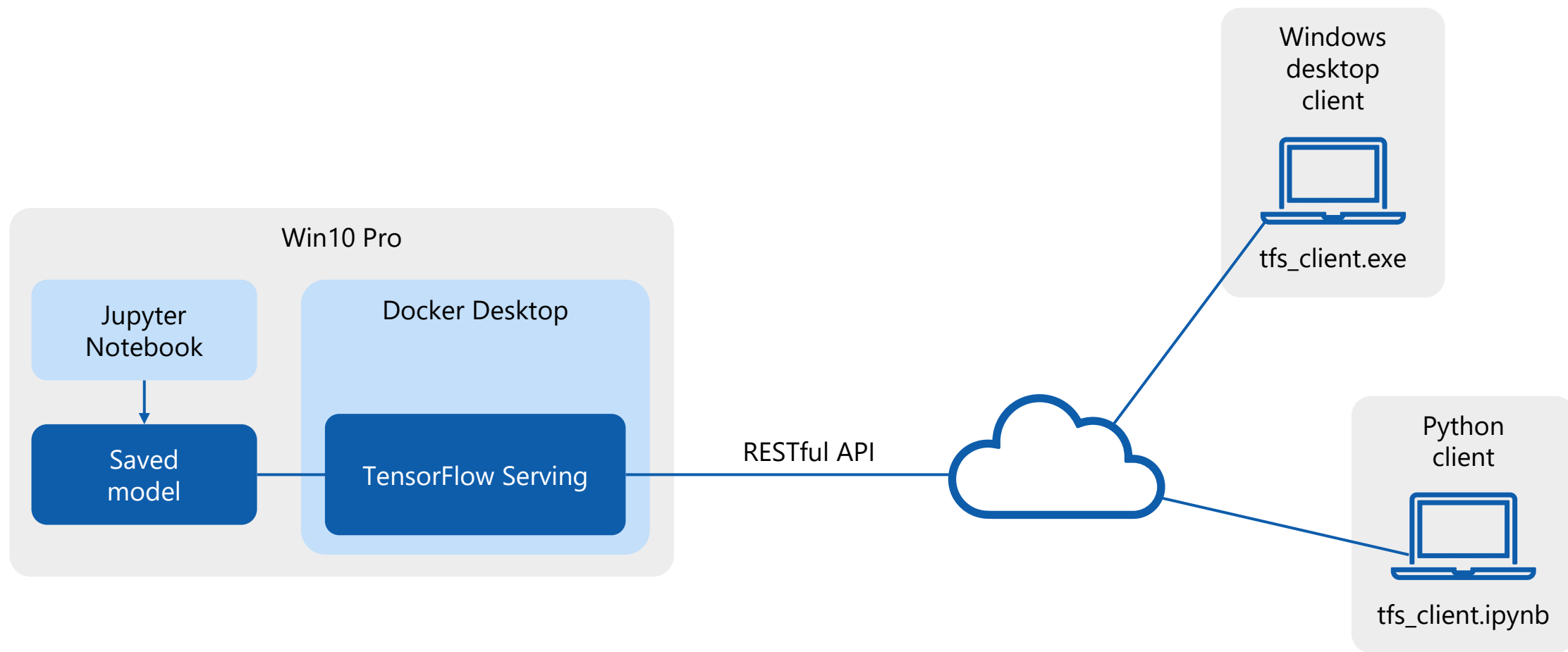
Оптимальные наборы признаков

Регрессор	Линейная регрессия	Ridge	Полиномиальная регрессия
Оптимальные наборы признаков	[5] [4 5] [3 4 5] [3 4 5 7] [3 4 5 6 7] [2 3 4 5 6 7]	[5] [4 5] [3 4 5] [3 4 5 7] [3 4 5 6 7] [1 3 4 5 6 7]	[5] [4 5] [3 4 5] [3 4 5 10] [2 3 4 5 10] [2 3 4 5 9 10] [2 3 4 5 6 9 10] [1 2 3 4 5 6 9 10] [1 2 3 4 5 6 8 9 10]

Регрессор	RandomForest	kNN	Градиентный бустинг
Оптимальные наборы признаков	[1] [1 6] [1 5 6] [1 3 5 6] [1 2 3 5 6] [1 2 3 4 5 6]	[1] [1 6] [1 5 6] [1 4 5 6] [1 2 4 5 6] [1 2 3 4 5 6]	[1] [1 6] [1 5 6] [1 4 5 6] [1 3 4 5 6] [1 2 3 4 5 6]



Схема приложения





ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru