



**ОБРАЗОВАТЕЛЬНЫЙ  
ЦЕНТР** МГТУ им. Н.Э.Баумана

# ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

## Прогнозирование конечных свойств новых материалов (композиционных материалов)

Докладчик: **Кислицин Никита Сергеевич**

слушатель Образовательного центра МГТУ им. Н. Э. Баумана

# Цель работы:

Создание прогнозных моделей ряда конечных свойств композиционных материалов на основе набора входных параметров и сокращение количества проводимых физических испытаний.

## Задачи:

- Изучение теоретических основ и методов решения ;
- Разведочный анализ входных данных;
- Предобработка данных;
- Обучение нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении. При построении модели необходимо 30% данных оставить на тестирование модели, на остальных происходит обучение моделей;
- Написание нейронной сети, которая будет рекомендовать соотношение матрица-наполнитель;
- Разработка приложения Flask, которое будет выдавать прогноз;
- Оценка точности модели на тренировочном и тестовом датасете;
- Создание репозитория в GitHub и размещение в нем кода исследования.

# Методы решения задачи:

1. Обучение моделей для прогноза модуля упругости при растяжении и прочности при растяжении с использованием библиотеки Scikit-learn следующими методами:

- линейная регрессия (linear regression),
- случайный лес (random forest),
- К-ближайших соседей (k-nearest neighbors),
- градиентный бустинг (gradient boosting),
- многослойный перцептрон (MLP).

2. Создание нейронных сетей с использованием модуля TensorFlow Keras для построения и тренировки моделей глубокого обучения, которая будет рекомендовать соотношение матрица-наполнитель.

# Входные данные начальных свойств компонентов

Базальтопластик (X\_br):

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, C_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2
0.0	1.857143	2030.000000	738.736842	30.000000	22.267857	100.000000	210.000000	70.000000	3000.000000	220.000000
1.0	1.857143	2030.000000	738.736842	50.000000	23.750000	284.615385	210.000000	70.000000	3000.000000	220.000000
2.0	1.857143	2030.000000	738.736842	49.900000	33.000000	284.615385	210.000000	70.000000	3000.000000	220.000000
3.0	1.857143	2030.000000	738.736842	129.000000	21.250000	300.000000	210.000000	70.000000	3000.000000	220.000000
4.0	2.771331	2030.000000	753.000000	111.860000	22.267857	284.615385	210.000000	70.000000	3000.000000	220.000000
...	...	...	...	...	...	...	...	...	...	...
1018.0	2.271346	1952.087902	912.855545	86.992183	20.123249	324.774576	209.198700	73.090961	2387.292495	125.007669
1019.0	3.444022	2050.089171	444.732634	145.981978	19.599769	254.215401	350.660830	72.920827	2360.392784	117.730099
1020.0	3.280604	1972.372865	416.836524	110.533477	23.957502	248.423047	740.142791	74.734344	2662.906040	236.606764
1021.0	3.705351	2066.799773	741.475517	141.397963	19.246945	275.779840	641.468152	74.042708	2071.715856	197.126067
1022.0	3.808020	1890.413468	417.316232	129.183416	27.474763	300.952708	758.747882	74.309704	2856.328932	194.754342

1023 rows × 10 columns

Нашивки из углепластика (X\_nup):

	Угол нашивки, град	Шаг нашивки	Плотность нашивки
0.0	0.0	4.000000	57.000000
1.0	0.0	4.000000	60.000000
2.0	0.0	4.000000	70.000000
3.0	0.0	5.000000	47.000000
4.0	0.0	5.000000	57.000000
...	...	...	...
1035.0	90.0	8.088111	47.759177
1036.0	90.0	7.619138	66.931932
1037.0	90.0	9.800926	72.858286
1038.0	90.0	10.079859	65.519479
1039.0	90.0	9.021043	66.920143

1040 rows × 3 columns

# Разведочный анализ данных

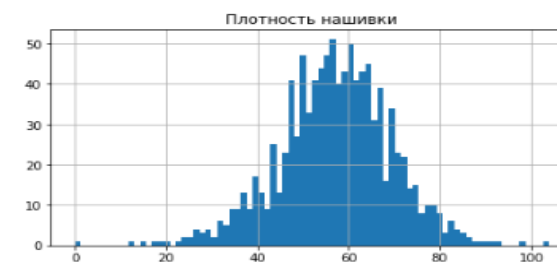
Описательная статистика по каждому параметру

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1023.0	2.93	0.91	0.39	2.32	2.91	3.55	5.59
Плотность, кг/м3	1023.0	1975.73	73.73	1731.76	1924.16	1977.62	2021.37	2207.77
модуль упругости, ГПа	1023.0	739.92	330.23	2.44	500.05	739.66	961.81	1911.54
Количество отвердителя, м.%	1023.0	110.57	28.30	17.74	92.44	110.56	129.73	198.95
Содержание эпоксидных групп,%_2	1023.0	22.24	2.41	14.25	20.61	22.23	23.96	33.00
Температура вспышки, С_2	1023.0	285.88	40.94	100.00	259.07	285.90	313.00	413.27
Поверхностная плотность, г/м2	1023.0	482.73	281.31	0.60	266.82	451.86	693.23	1399.54
Модуль упругости при растяжении, ГПа	1023.0	73.33	3.12	64.05	71.25	73.27	75.36	82.68
Прочность при растяжении, МПа	1023.0	2466.92	485.63	1036.86	2135.85	2459.52	2767.19	3848.44
Потребление смолы, г/м2	1023.0	218.42	59.74	33.80	179.63	219.20	257.48	414.59
Угол нашивки, град	1023.0	44.25	45.02	0.00	0.00	0.00	90.00	90.00
Шаг нашивки	1023.0	6.90	2.56	0.00	5.08	6.92	8.59	14.44
Плотность нашивки	1023.0	57.15	12.35	0.00	49.80	57.34	64.94	103.99

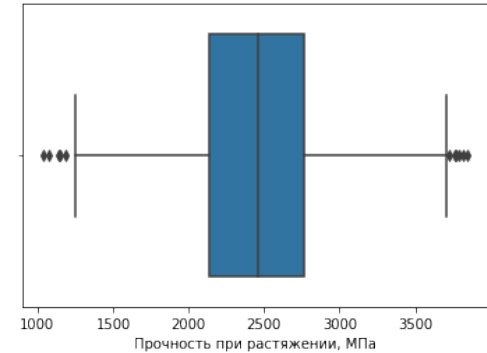
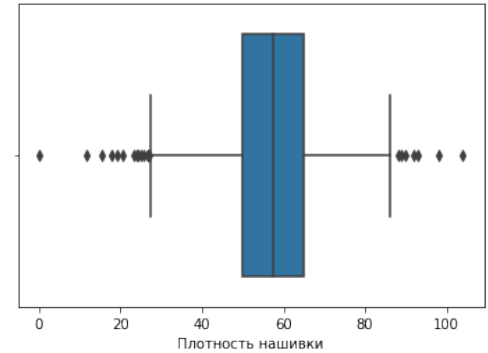
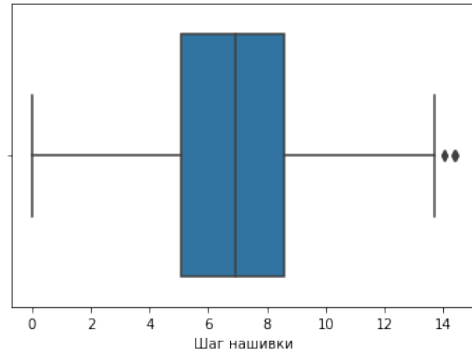
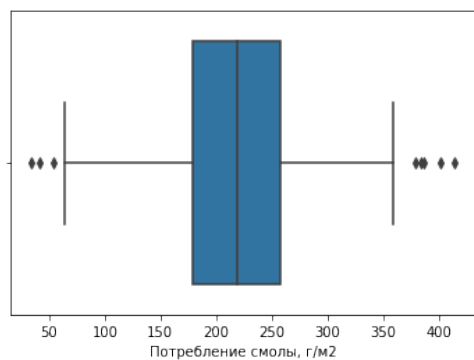
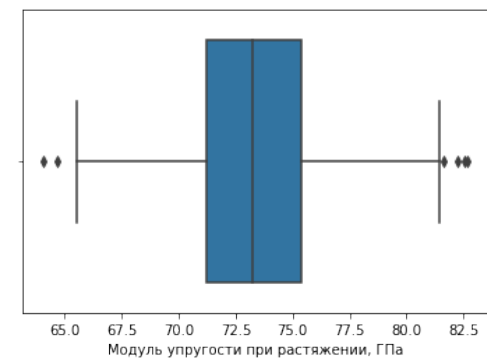
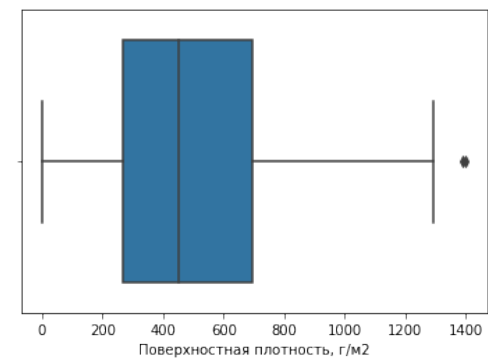
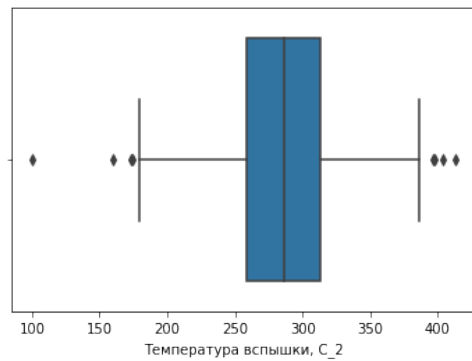
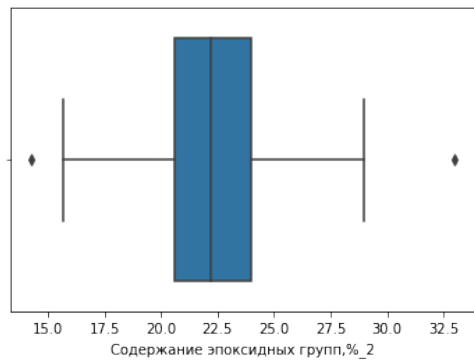
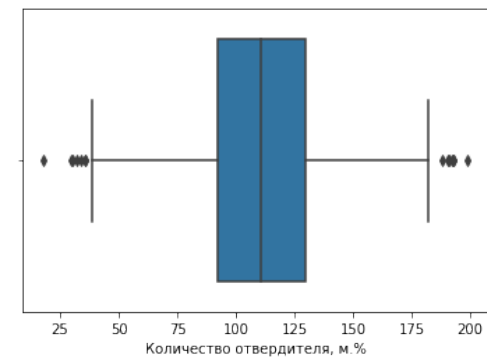
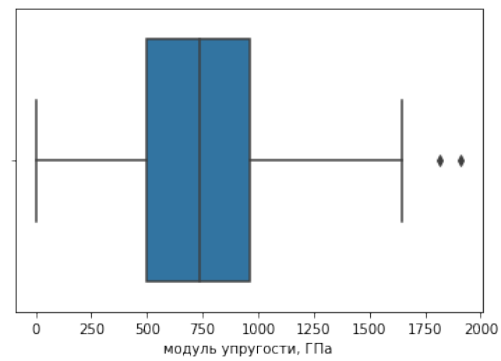
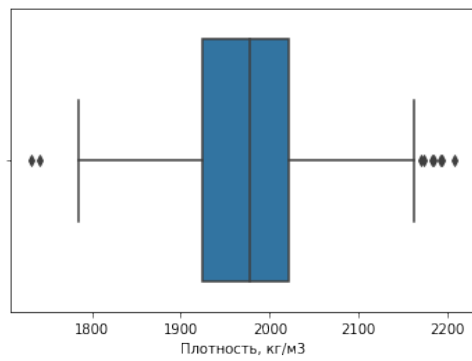
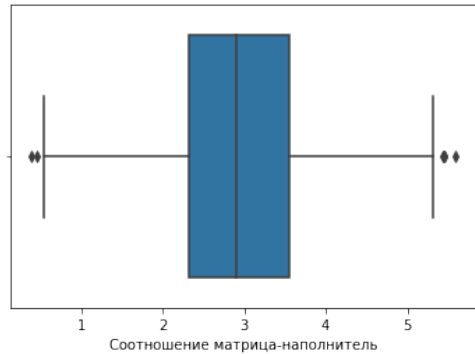
Кол-во уникальных значений

Соотношение матрица-наполнитель	1014
Плотность, кг/м3	1013
модуль упругости, ГПа	1020
Количество отвердителя, м.%	1005
Содержание эпоксидных групп,%_2	1004
Температура вспышки, С_2	1003
Поверхностная плотность, г/м2	1004
Модуль упругости при растяжении, ГПа	1004
Прочность при растяжении, МПа	1004
Потребление смолы, г/м2	1003
Угол нашивки, град	2
Шаг нашивки	989
Плотность нашивки	988
dtype: int64	

# Гистограммы распределения

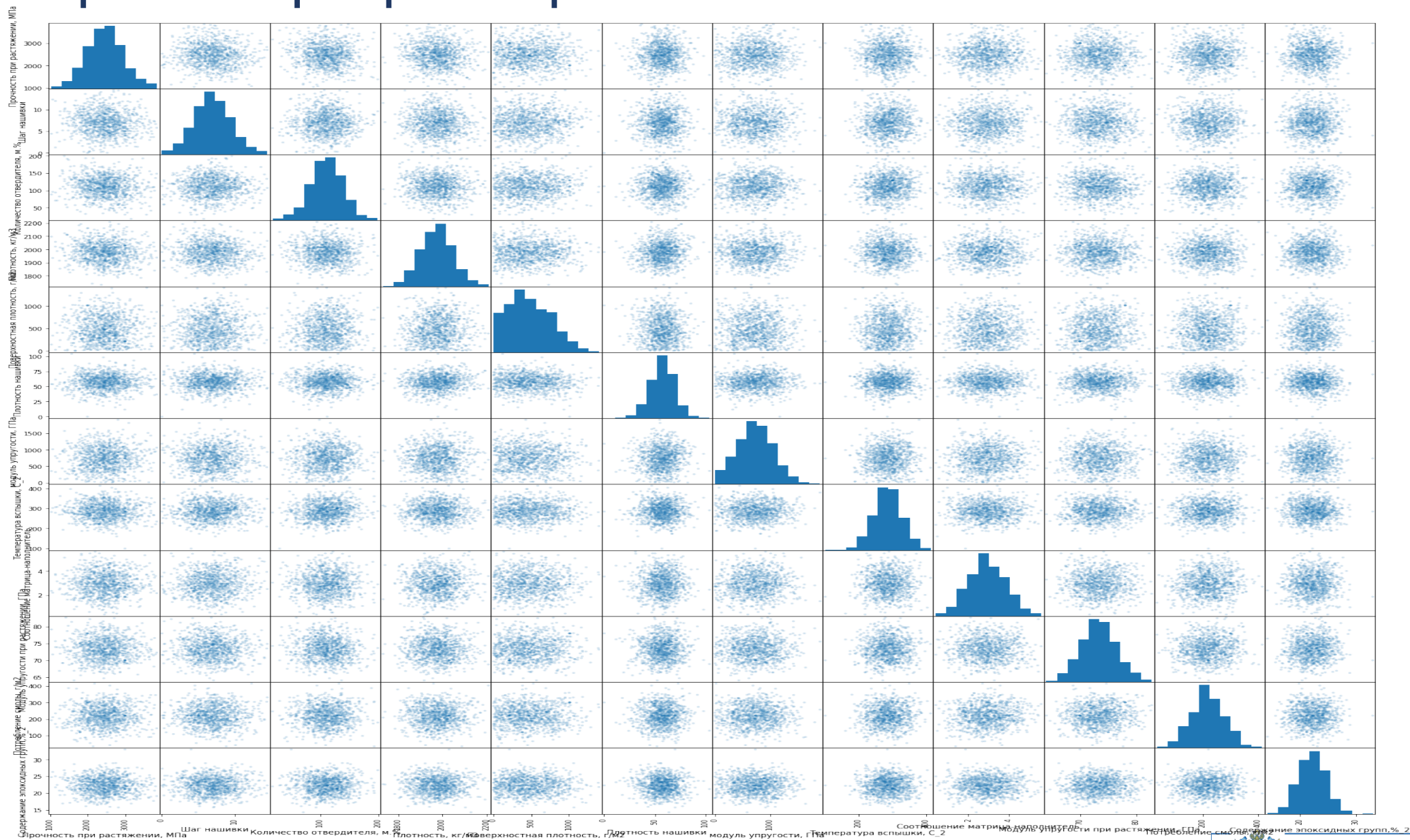


# Диаграммы размаха





# Попарные графики рассеяния





# Предобработка данных

## Исключение выбросов методом на основе межквартильных диапазонов

```
for column in df.columns:
    q75,q25 = np.percentile(df.loc[:,column],[75,25])
    iqr = q75-q25

    max = q75+(1.5*iqr)
    min = q25-(1.5*iqr)

    df.loc[df[column] < min,column] = np.nan
    df.loc[df[column] > max,column] = np.nan
```

## Нормализация данных

```
# масштабирование значений в диапазон [0, 1]
scaler = MinMaxScaler()
df_norm=pd.DataFrame(scaler.fit_transform(df), columns=df.columns, index=df.index)
```

```
df_norm.head(5)
```

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, С_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашив
1.0	0.274768	0.651097	0.447061	0.079153	0.607435	0.509164	0.16223	0.280303	0.71259	0.529221	0.0	0.289334	0.5571
3.0	0.274768	0.651097	0.447061	0.630983	0.418887	0.583596	0.16223	0.280303	0.71259	0.529221	0.0	0.362355	0.3358
4.0	0.466552	0.651097	0.455721	0.511257	0.495653	0.509164	0.16223	0.280303	0.71259	0.529221	0.0	0.362355	0.5060
5.0	0.465836	0.571539	0.452685	0.511257	0.495653	0.509164	0.16223	0.280303	0.71259	0.529221	0.0	0.362355	0.5571
6.0	0.424236	0.332865	0.488508	0.511257	0.495653	0.509164	0.16223	0.280303	0.71259	0.529221	0.0	0.362355	0.7273

# Результат препроцессинга

Тепловая карта после нормализации



# Обучение моделей для прогноза

## Метрики качества моделей

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

Where,

$\hat{y}$  – predicted value of  $y$

$\bar{y}$  – mean value of  $y$

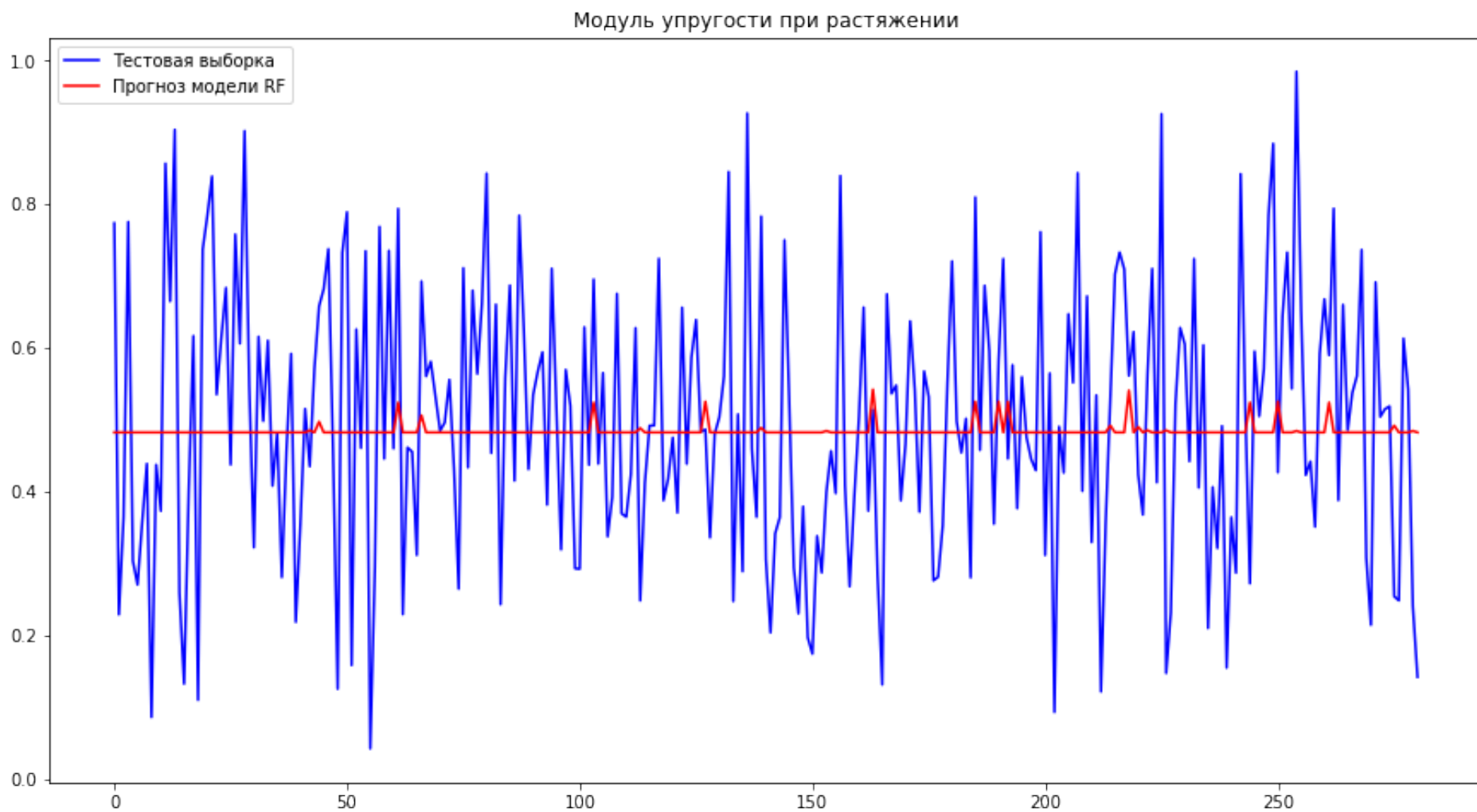
```
'MSE': metrics.mean_squared_error(y_test_upr, y_pred_upr),  
'MAE': metrics.mean_absolute_error(y_test_upr, y_pred_upr),  
'R^2': metrics.r2_score(y_test_upr, y_pred_upr)
```

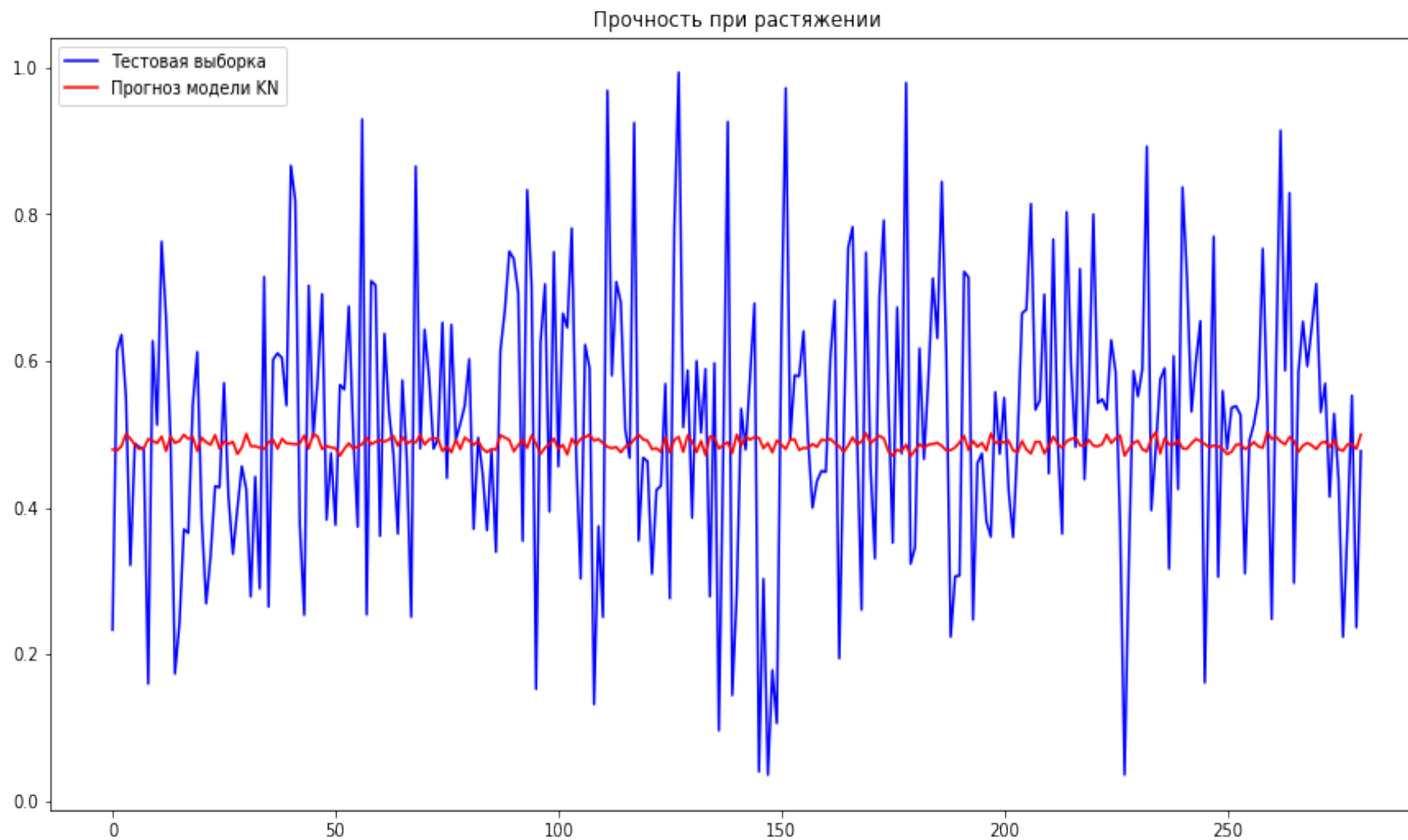
## Модуль упругости при растяжении

Model	MSE	MAE	R^2
DummyRegression_mean	0.034832	0.151202	-0.005930
MLPRegressor_best	0.034835	0.151209	-0.005994
MLPRegressor_default	0.036425	0.156268	-0.051938
GradientBooster_best	0.035135	0.151845	-0.014684
GradientBooster_default	0.037769	0.155943	-0.090727
KNeighborsRegressor_best	0.034724	0.151232	-0.002795
RandomForestRegressor_best	0.034667	0.150828	-0.001151
KNeighborsRegressor_default	0.041694	0.166480	-0.204103
LinearRegression_best	0.034808	0.152207	-0.005222
LinearRegression_default	0.034808	0.152207	-0.005222
RandomForestRegressor_default	0.036049	0.154849	-0.041076

## Прочность при растяжении

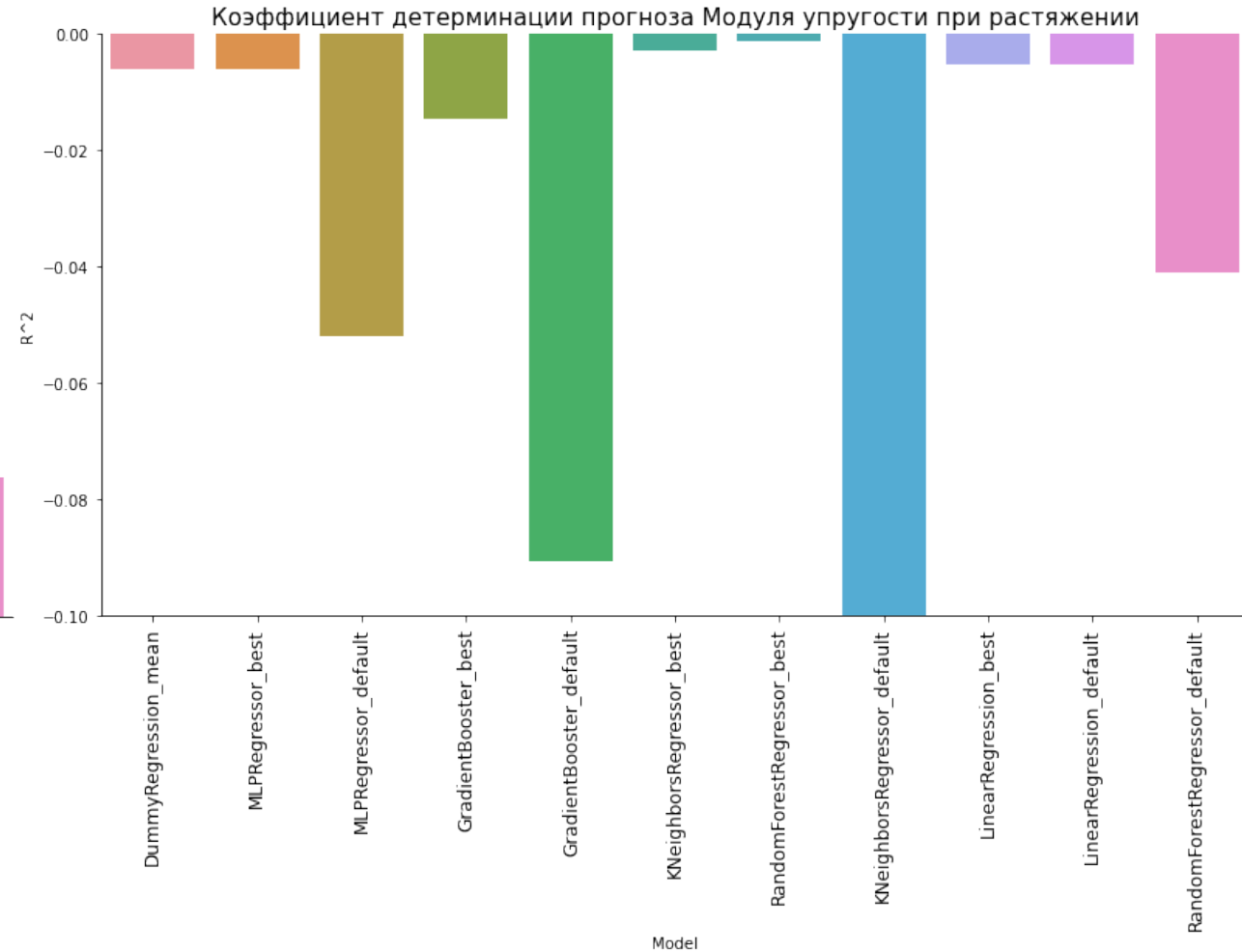
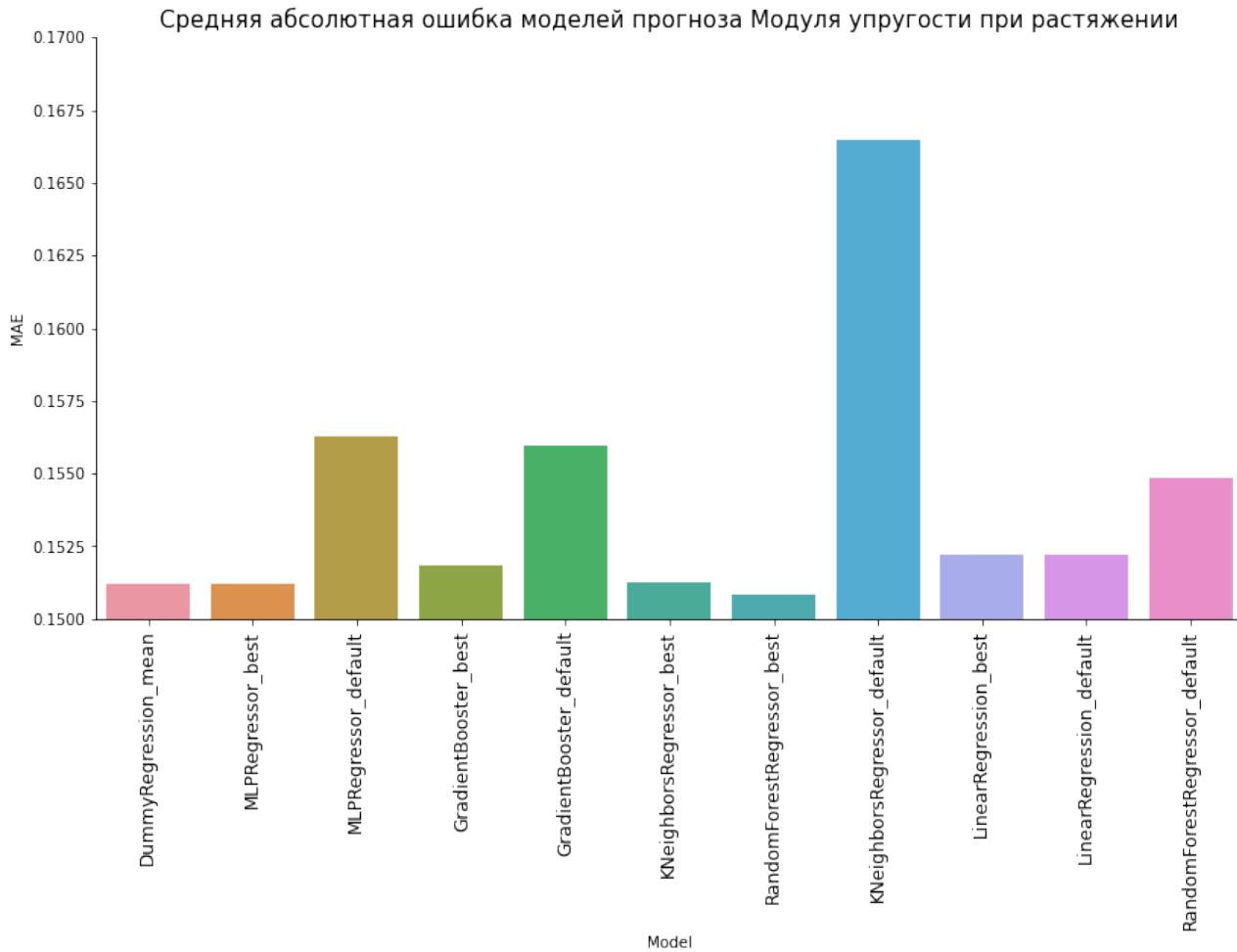
Model	MSE	MAE	R^2
RandomForestRegressor_default	0.035997	0.152083	-0.063189
RandomForestRegressor_best	0.035039	0.147426	-0.034884
MLPRegressor_default	0.037476	0.154898	-0.106881
LinearRegression_best	0.035449	0.147903	-0.046999
KNeighborsRegressor_default	0.041427	0.157250	-0.223582
KNeighborsRegressor_best	0.034805	0.146584	-0.027995
LinearRegression_default	0.035449	0.147903	-0.046999
GradientBooster_default	0.037476	0.154898	-0.106881
GradientBooster_best	0.035680	0.150019	-0.053820
DummyRegression_mean	0.034862	0.146748	-0.029672
MLPRegressor_best	0.035616	0.148546	-0.051939



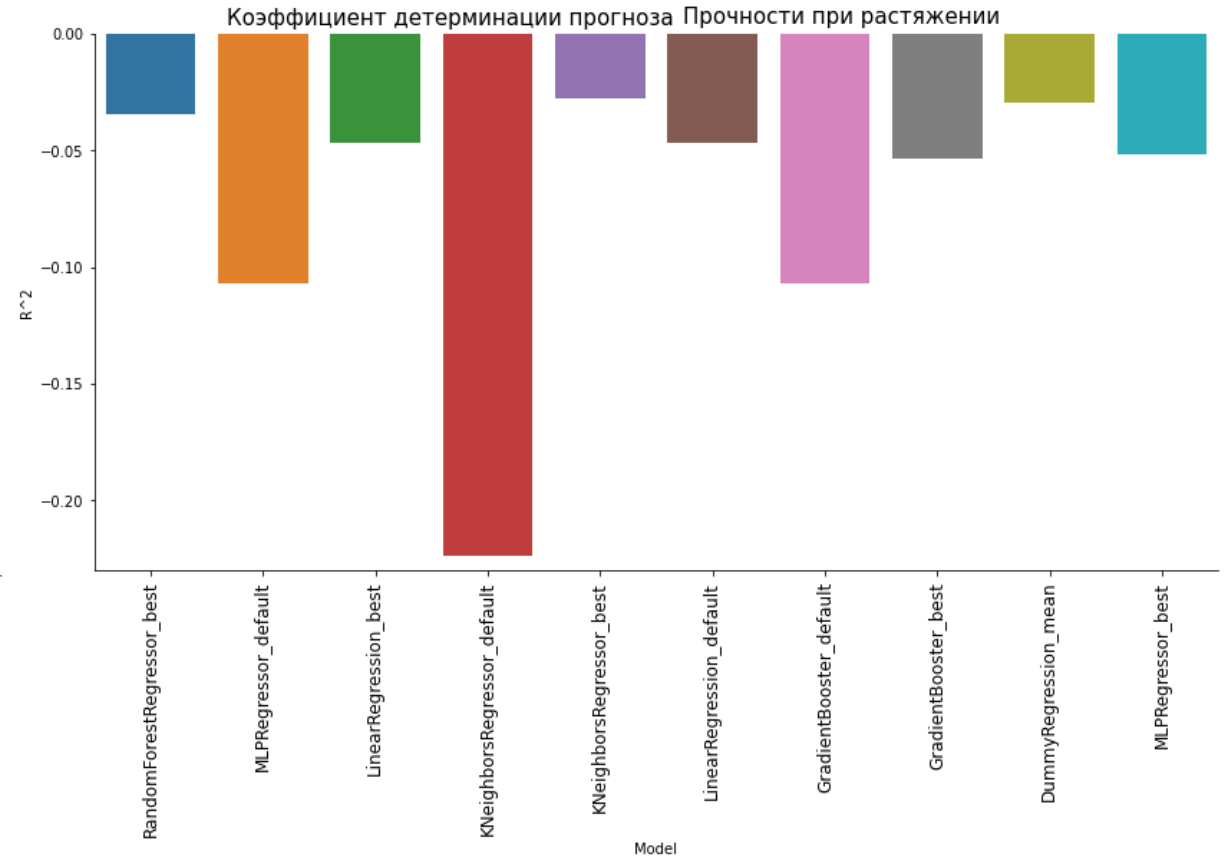
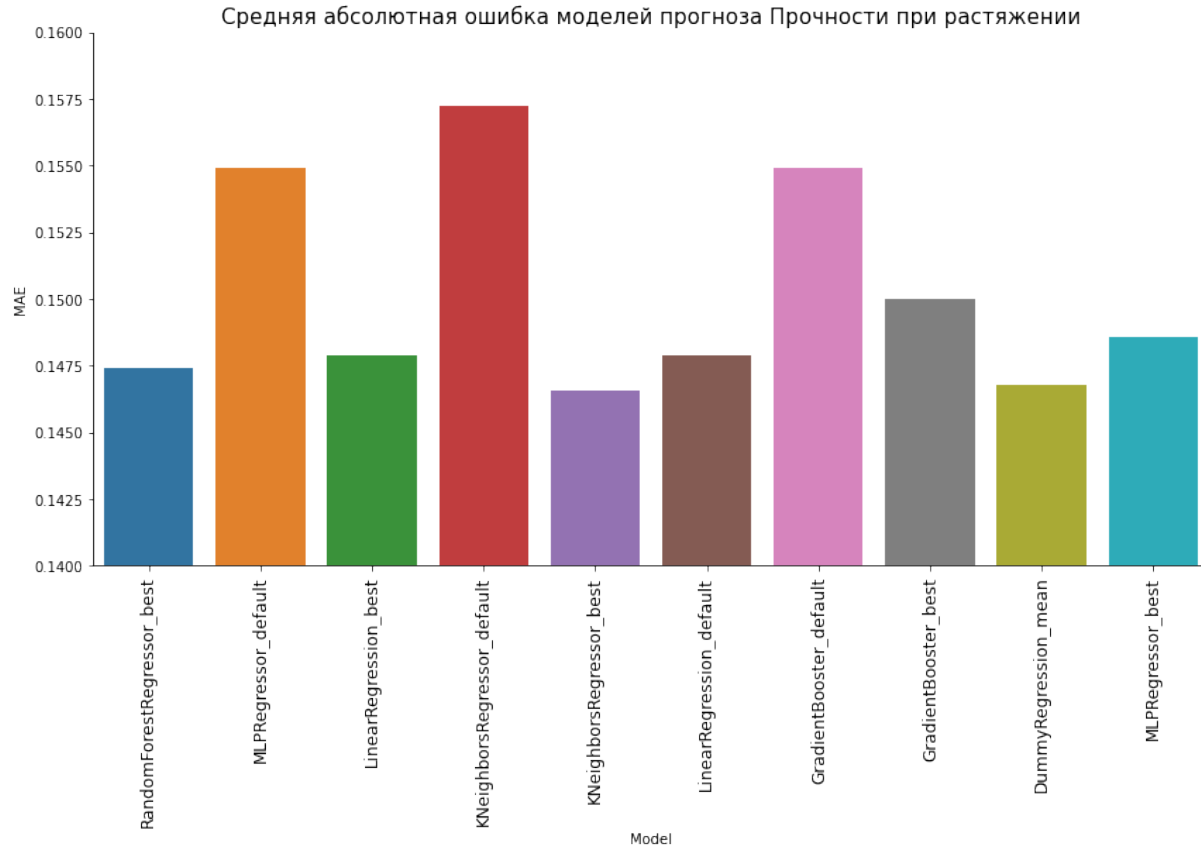




# Диаграммы метрик прогноза модуля упругости при растяжении

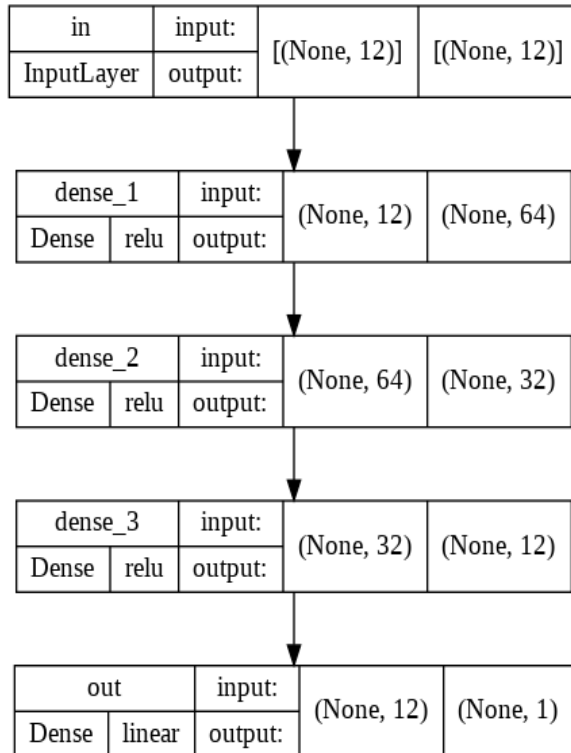


# Диаграммы метрик прогноза прочности при растяжении



# Нейронные сети для соотношения матрица-наполнитель

## Архитектура нейросети

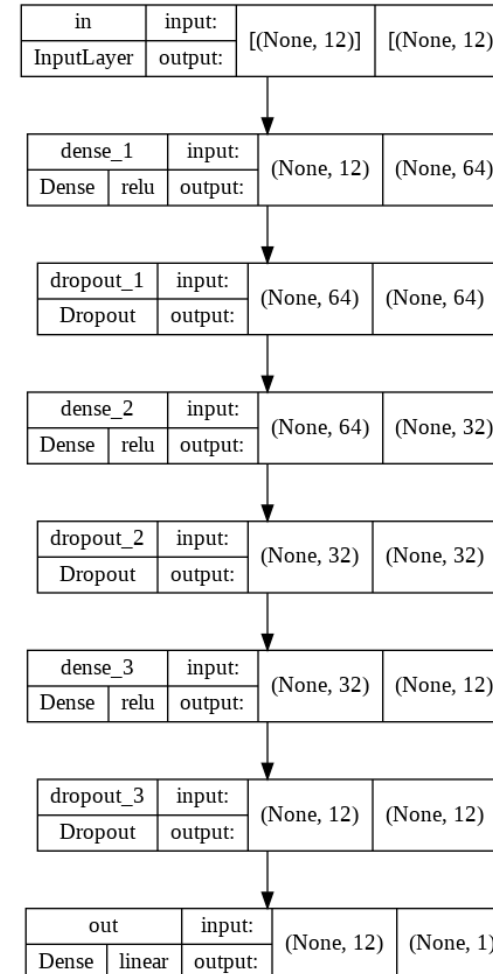


Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
dense_1 (Dense)	(None, 64)	832
dense_2 (Dense)	(None, 32)	2080
dense_3 (Dense)	(None, 12)	396
out (Dense)	(None, 1)	13

Total params: 3,321  
Trainable params: 3,321  
Non-trainable params: 0

## Архитектура нейросети с dropout



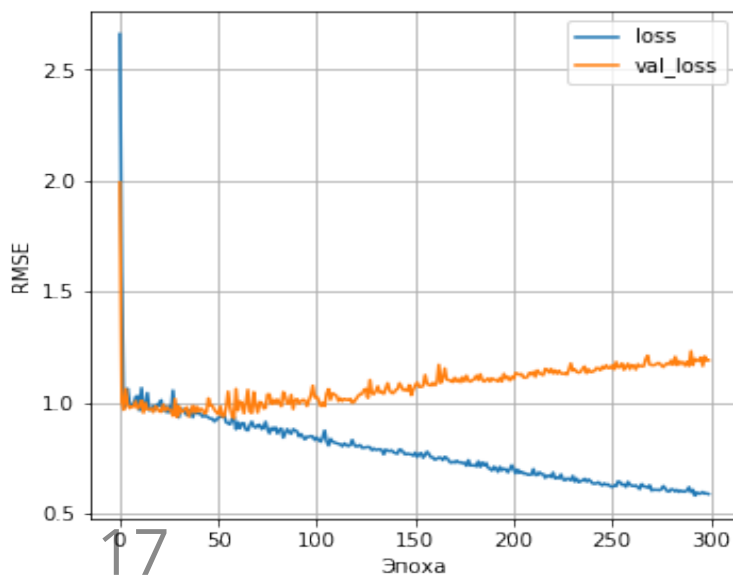
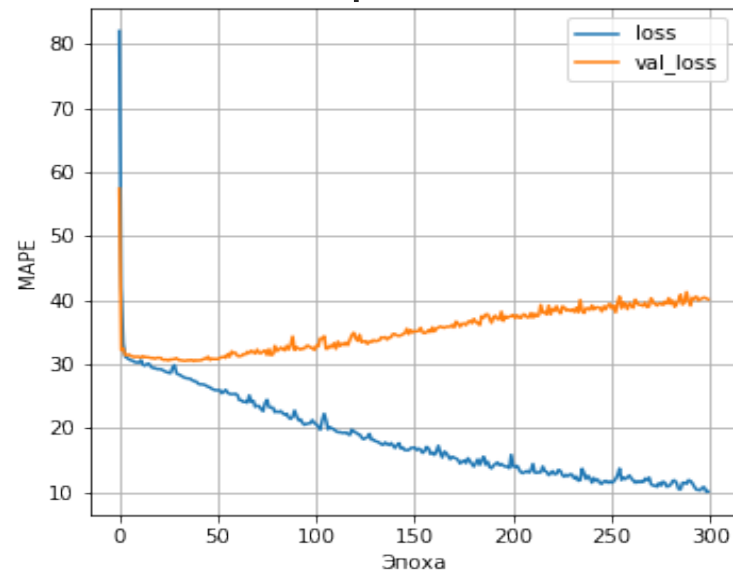
Model: "sequential\_2"

Layer (type)	Output Shape	Param #
=====		
dense_1 (Dense)	(None, 64)	832
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 32)	2080
dropout_2 (Dropout)	(None, 32)	0
dense_3 (Dense)	(None, 12)	396
dropout_3 (Dropout)	(None, 12)	0
out (Dense)	(None, 1)	13

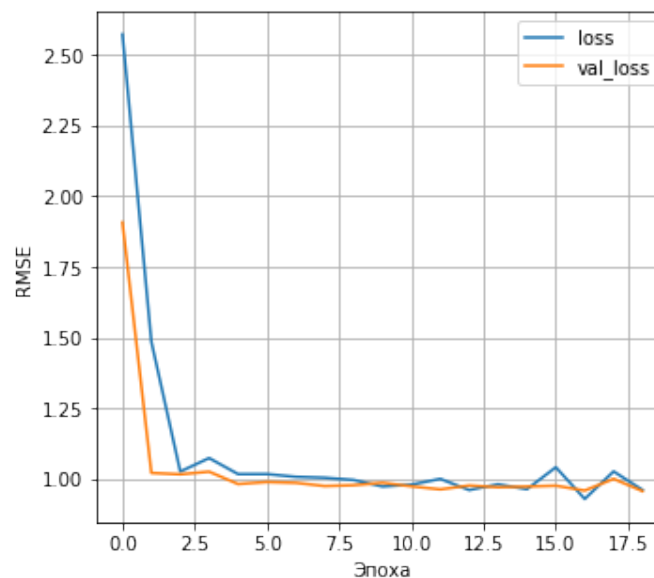
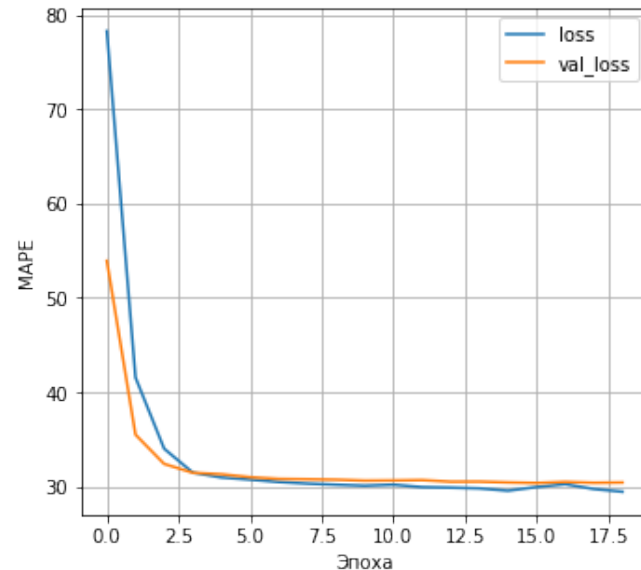
Total params: 3,321  
Trainable params: 3,321  
Non-trainable params: 0

# Метрики нейронных сетей

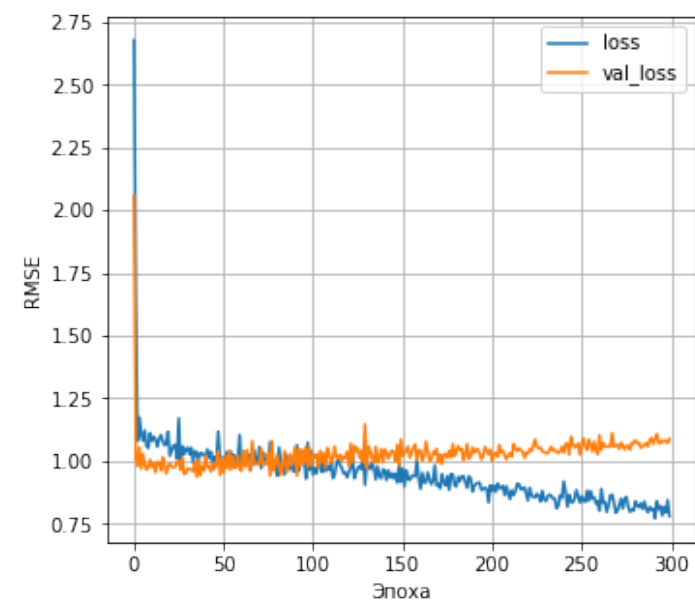
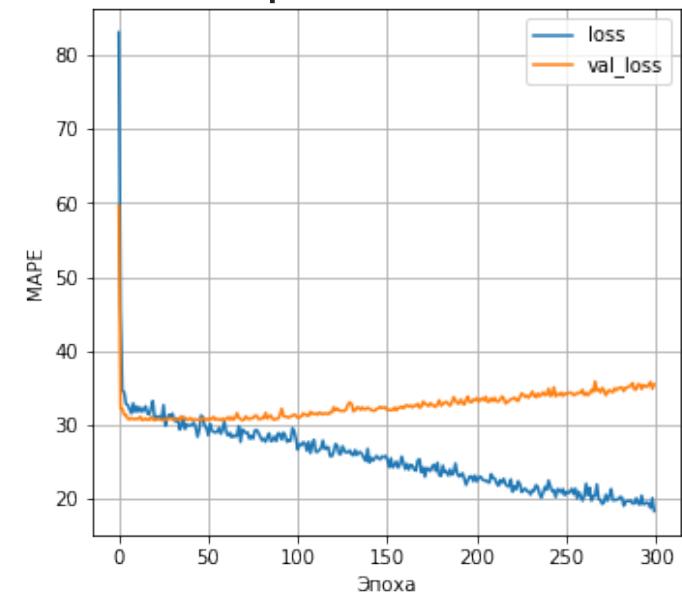
## ■ Стандарт

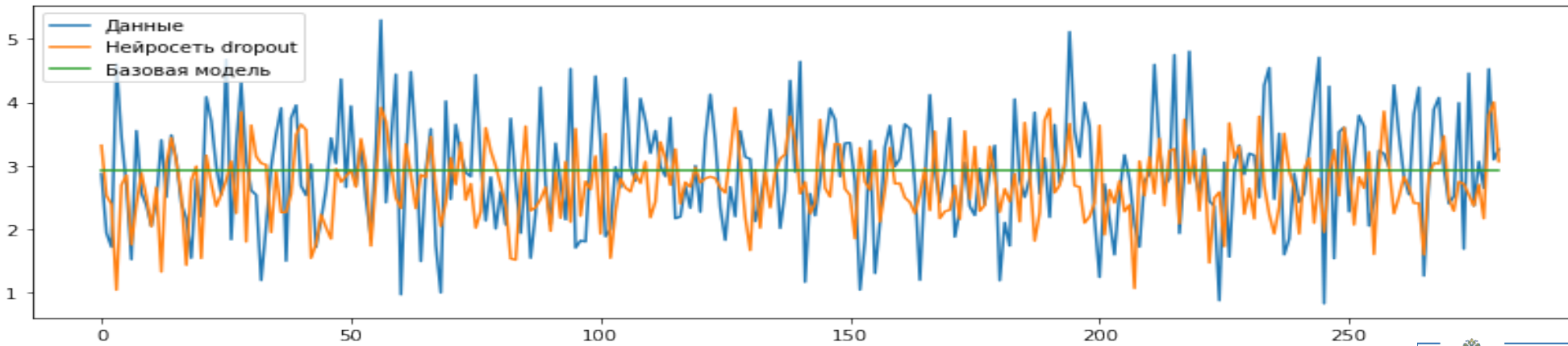
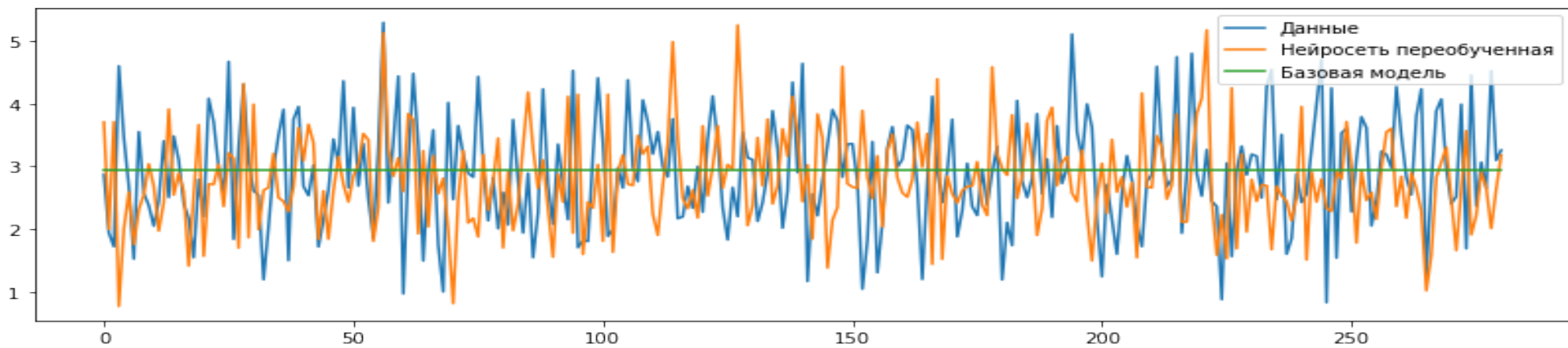


## ■ Ранняя остановка



## ■ Dropout







# Оценка моделей на тренировочной и тестовой выборках

## Метрики нейронных сетей

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.001569	-0.873433	-0.694175	-0.298186	-2.359781
Нейросеть переобученная	-0.778836	-1.164011	-0.928316	-0.368104	-3.836049
Нейросеть с ранней остановкой	-0.127695	-0.926798	-0.751660	-0.282948	-2.410464
Нейросеть dropout	-0.381299	-1.025730	-0.816692	-0.316429	-3.561940

	R2	RMSE	MAE	MAPE	max_error
Модуль упругости при растяжении, тренировочный	0.011929	-0.192269	-0.154746	-3.382084e+12	-0.517673
Модуль упругости при растяжении, тестовый	-0.001151	-0.186190	-0.150828	-4.486452e-01	-0.500318

Случайный лес

Метод К ближайших соседей

Нейронная сеть с ранней остановкой

	R2	RMSE	MAE	MAPE	max_error
Прочность при растяжении, тренировочный	0.006495	-0.189430	-0.151106	-3.387012e+12	-0.508566
Прочность при растяжении, тестовый	-0.027995	-0.186562	-0.146584	-4.724052e-01	-0.496947

	R2	RMSE	MAE	MAPE	max_error
Соотношение матрица-наполнитель, тренировочный	-0.116392	-0.952711	-0.76551	-0.294564	-2.989776
Соотношение матрица-наполнитель, тестовый	-0.127695	-0.926798	-0.75166	-0.282948	-2.410464

# Flask приложение для профильных специалистов

## Прогнозирование модуля упругости при растяжении и прочности при растяжении

Соотношение матрица-наполнитель (0..6)

Плотность, кг/м3 (1700...2300)

Модуль упругости, ГПа (2...2000)

Количество отвердителя, м.% (17...200)

Содержание эпоксидных групп,%\_2 (14...34)

Температура вспышки, C\_2 (100...414)

Поверхностная плотность, г/м2 (0.6...1400)

Потребление смолы, г/м2 (33...414)

Угол нашивки, град (0 или 90)

Шаг нашивки (0...15)

Плотность нашивки (0...104)

Входные переменные:

	Соотношение матрица-наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, C_2	Поверхностная плотность, г/м2	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
0	4.029126	1880.0	622.0	111.86	22.267857	284.615385	470.0	220.0	90.0	4.0	60.0

Результат модели:

Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа
72.81891497929365	2523.9223070281537

# Flask приложение для профильных специалистов

## Прогнозирование соотношения матрица-наполнитель

Плотность, кг/м3 (1700...2300)

Модуль упругости, ГПа (2...2000)

Количество отвердителя, м.% (17...200)

Содержание эпоксидных групп,%\_2 (14...34)

Температура вспышки, C\_2 (100...414)

Поверхностная плотность, г/м2 (0.6...1400)

Модуль упругости при растяжении, ГПа (64...83)

Прочность при растяжении, МПа (1036...3849)

Потребление смолы, г/м2 (33...414)

Угол нашивки, град (0 или 90)

Шаг нашивки (0...15)

Плотность нашивки (0...104)

Входные переменные:

	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м.%	Содержание эпоксидных групп,%_2	Температура вспышки, C_2	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Угол нашивки, град	Шаг нашивки	Плотность нашивки
0	1880.0	622.0	111.86	22.267857	284.615385	470.0	73.333333	2455.555556	220.0	90.0	4.0	60.0

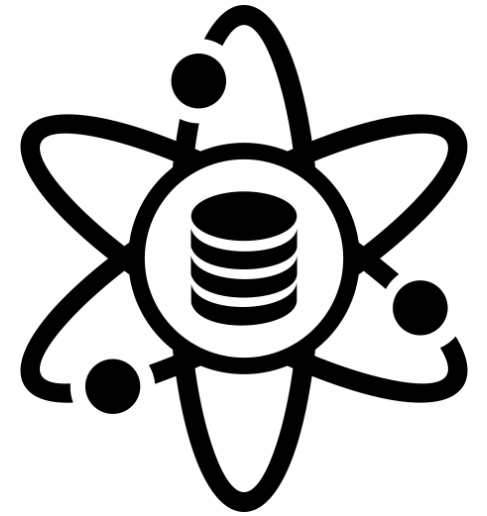
Результат модели:

Соотношение матрица-наполнитель
2.5154960585858928

# Заключение

По итогу выполнения данной квалификационной работы можно сделать следующие выводы:

- распределение исходных данных близко к нормальному;
- корреляция между парами признаков практически отсутствует;
- примененные модели регрессии и нейронных сетей не показали высокой эффективности в прогнозировании свойств композитов в этом конкретном исследовании, необходимы дополнительные вводные данные для улучшения моделей;
- В ходе выполнения данной работы был полностью проработан весь pipeline машинного обучения, рассмотрена большая область задач и операций, которые приходится выполнять специалисту по Data Science.





[edu.bmstu.ru](http://edu.bmstu.ru)

**+7 495 182-83-85**

[edu@bmstu.ru](mailto:edu@bmstu.ru)

Москва, Госпитальный переулок ,  
д. 4-6, с.3



**ОБРАЗОВАТЕЛЬНЫЙ  
ЦЕНТР** МГТУ им. Н. Э. Баумана