



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование конечных свойств новых материалов (композиционных материалов).

Никифоров Денис Николаевич



Этапы выполнения работы

- ❖ Разведочный анализ данных
- ❖ Предобработка данных
- ❖ Разработка и обучение регрессионных моделей для прогнозирования «Модуль упругости при растяжении, ГПа» и «Прочность при растяжении, МПа»
- ❖ Построение нейронной сети для рекомендации «Соотношение матрица-наполнитель»



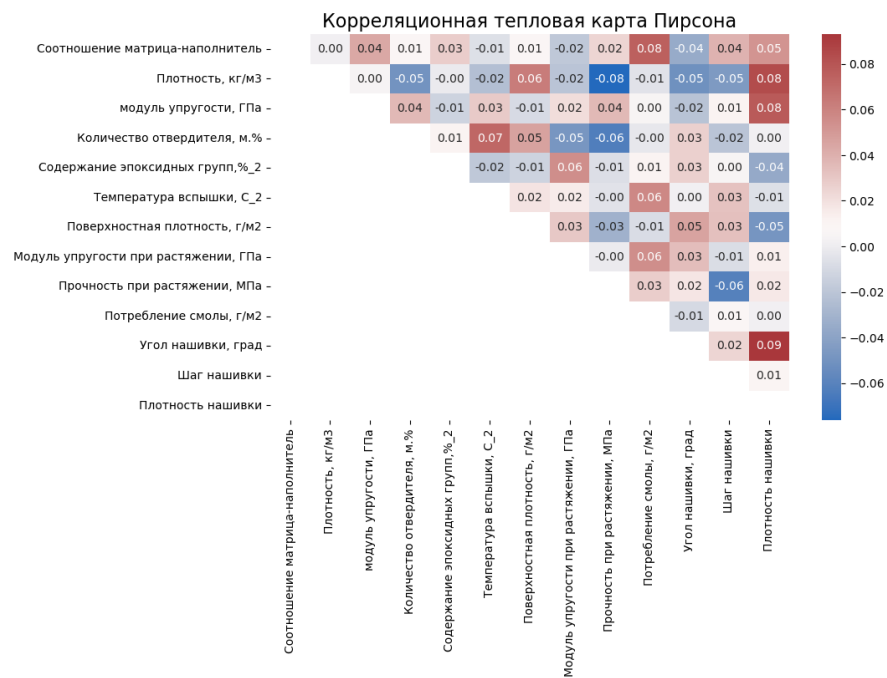
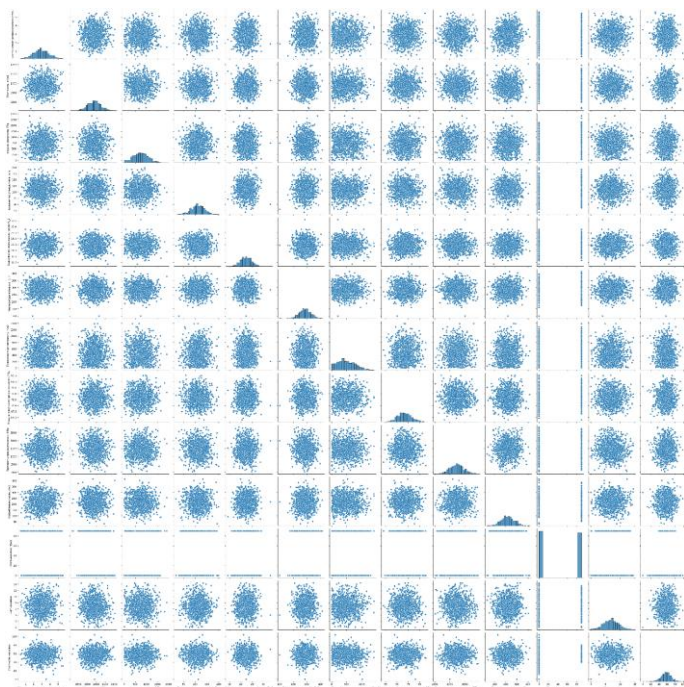
Разведочный анализ данных

Дополнительная информация

- описательная статистика данного датасета;
- визуальный анализ гистограмм;
- визуальный анализ диаграмм размаха («ящик с усами»);
- проверка нормальности распределения по критерию Пирсона;
- анализ попарных графиков рассеяния переменных;
- корреляционный анализ с целью поиска коэффициентов



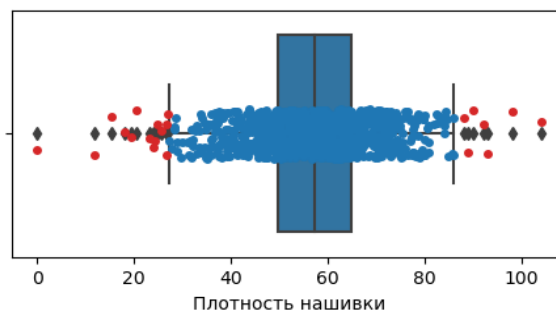
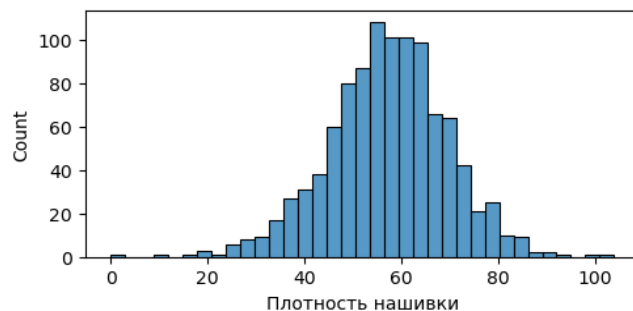
Разведочный анализ данных



Дополнительная информация



Предобработка данных



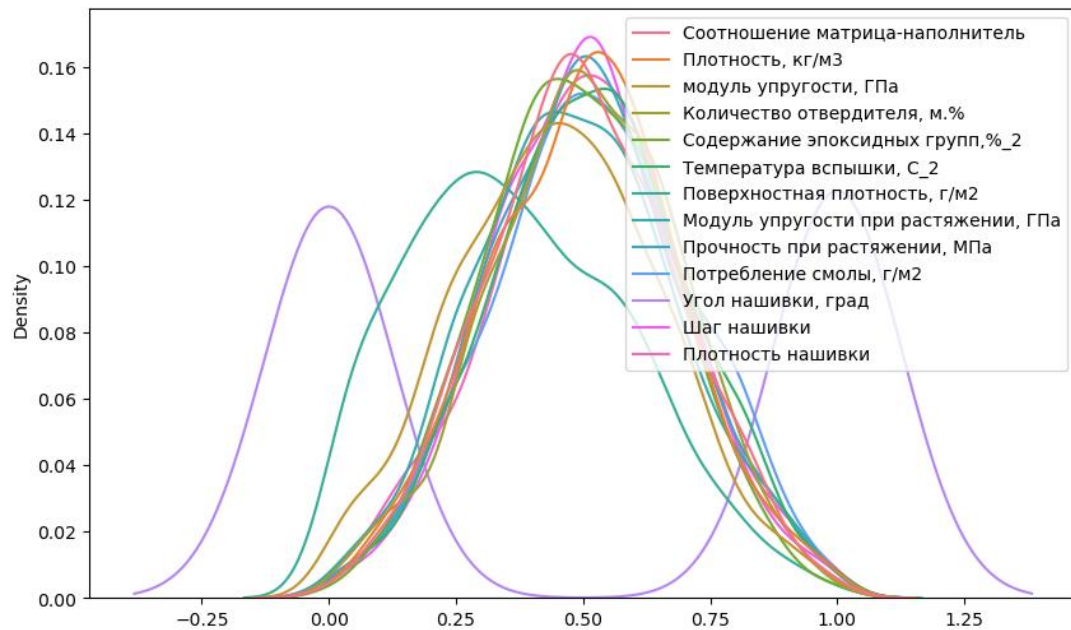
Дополнительная информация

- Удаление выбросов методом межквартильных расстояний

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1023.0	2.930366	0.913222	0.389403	2.317887	2.906878	3.552660	5.591742
Плотность, кг/м3	1023.0	1975.734888	73.729231	1731.764635	1924.155467	1977.621657	2021.374375	2207.773481
модуль упругости, ГПа	1023.0	739.923233	330.231581	2.436909	500.047452	739.664328	961.812526	1911.536477
Количество отвердителя, м.%	1023.0	110.570769	28.295911	17.740275	92.443497	110.564840	129.730366	198.953207
Содержание эпоксидных групп, %_2	1023.0	22.244390	2.406301	14.254985	20.608034	22.230744	23.961934	33.000000
Температура вспышки, С_2	1023.0	285.882151	40.943260	100.000000	259.066528	285.896812	313.002106	413.273418
Поверхностная плотность, г/м2	1023.0	482.731833	281.314690	0.603740	266.816645	451.864365	693.225017	1399.542362
Модуль упругости при растяжении, ГПа	1023.0	73.328571	3.118983	64.054061	71.245018	73.268805	75.356612	82.682051
Прочность при растяжении, МПа	1023.0	2466.922843	485.628006	1036.856605	2135.850448	2459.524526	2767.193119	3848.436732
Потребление смолы, г/м2	1023.0	218.423144	59.735931	33.803026	179.627520	219.198882	257.481724	414.590628
Угол нашивки, град	1023.0	44.252199	45.015793	0.000000	0.000000	0.000000	90.000000	90.000000
Шаг нашивки	1023.0	6.899222	2.563467	0.000000	5.080033	6.916144	8.586293	14.440522
Плотность нашивки	1023.0	57.153929	12.350969	0.000000	49.799212	57.341920	64.944961	103.988901



Предобработка данных



	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	926.0	0.499687	0.187841	0.0	0.373077	0.494936	0.629774	1.0
Плотность, кг/м3	926.0	0.503041	0.188261	0.0	0.368184	0.511740	0.625266	1.0
модуль упругости, ГПа	926.0	0.451888	0.201469	0.0	0.305669	0.452599	0.587461	1.0
Количество отвердителя, м.%	926.0	0.506559	0.187611	0.0	0.378514	0.506532	0.639120	1.0
Содержание эпоксидных групп, %_2	926.0	0.491063	0.180438	0.0	0.367101	0.488912	0.623296	1.0
Температура вспышки, С_2	926.0	0.516443	0.190857	0.0	0.385988	0.516931	0.646553	1.0
Поверхностная плотность, г/м2	926.0	0.373626	0.216945	0.0	0.204863	0.356181	0.538397	1.0
Модуль упругости при растяжении, ГПа	926.0	0.487330	0.196140	0.0	0.353512	0.483718	0.617568	1.0
Прочность при растяжении, МПа	926.0	0.504018	0.189451	0.0	0.373350	0.501053	0.623037	1.0
Потребление смолы, г/м2	926.0	0.521132	0.194829	0.0	0.391647	0.523459	0.652734	1.0
Угол нашивки, град	926.0	0.509719	0.500176	0.0	0.000000	1.000000	1.000000	1.0
Шаг нашивки	926.0	0.503077	0.183709	0.0	0.372377	0.506414	0.626112	1.0
Плотность нашивки	926.0	0.505927	0.193315	0.0	0.379888	0.507146	0.631896	1.0



Обучение регрессионных моделей

Модуль упругости при растяжении, ГПа и Прочность при растяжении, МПа

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.014369	-0.191387	-0.156417	-0.609281	-0.451870
LinearRegression	-0.032590	-0.193068	-0.158108	-0.610380	-0.453773
Ridge	-0.030940	-0.192918	-0.158001	-0.610288	-0.453167
Lasso	-0.014369	-0.191387	-0.156417	-0.609281	-0.451870
SVR	-0.358672	-0.220606	-0.178406	-0.676160	-0.572981
KNeighborsRegressor	-0.310141	-0.216813	-0.177070	-0.681401	-0.522064
DecisionTreeRegressor	-1.138144	-0.274442	-0.219973	-0.753242	-0.696741
RandomForestRegressor	-0.083913	-0.197681	-0.161187	-0.633771	-0.481651

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.019621	-0.190403	-0.151815	-3521152612319.554688	-0.469354
LinearRegression	-0.033334	-0.191600	-0.153487	-3927564128168.418945	-0.469799
Ridge	-0.031568	-0.191445	-0.153329	-3908985374430.334473	-0.469586
Lasso	-0.019621	-0.190403	-0.151815	-3521152612319.554688	-0.469354
SVR	-0.299595	-0.213581	-0.171366	-3621535248060.870117	-0.510943
DecisionTreeRegressor	-1.055530	-0.269714	-0.216114	-4088487506854.995117	-0.689723
GradientBoostingRegressor	-0.140626	-0.200972	-0.161459	-4855337034318.583984	-0.504240

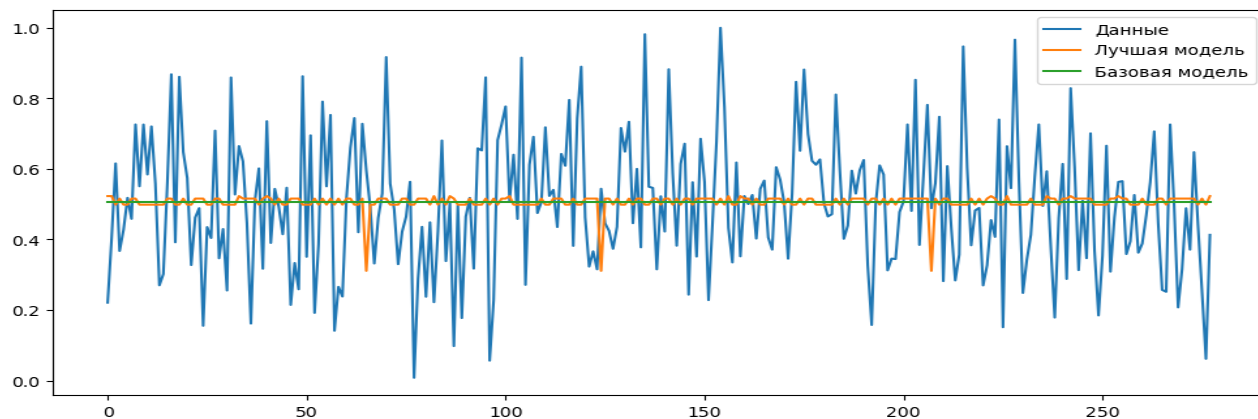
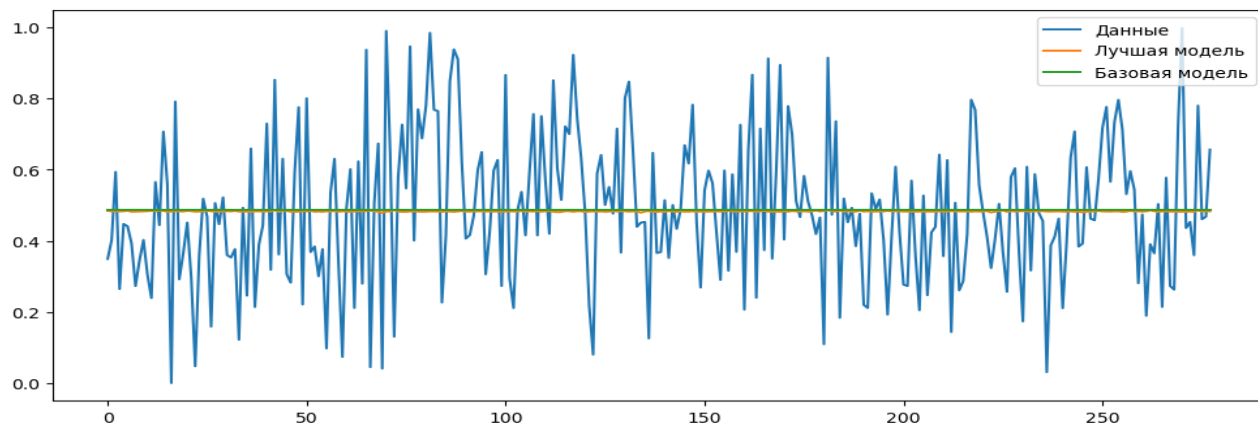
	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=1500, positive=True, solver='lbfgs')	-0.014343	-0.191386	-0.156439	-0.609219	-0.451881
Lasso(alpha=0.005)	-0.014369	-0.191387	-0.156417	-0.609281	-0.451870
SVR(C=0.04, kernel='sigmoid')	-0.013755	-0.191319	-0.156381	-0.604445	-0.452127
KNeighborsRegressor(n_neighbors=29)	-0.059535	-0.195578	-0.159547	-0.622749	-0.472972
DecisionTreeRegressor(criterion='poisson', max_depth=2, max_features=5, random_state=3000)	-0.019018	-0.191589	-0.155039	-0.607660	-0.460805

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=990, positive=True, solver='lbfgs')	-0.020010	-0.190438	-0.151845	-3520303507813.850586	-0.469313
Lasso(alpha=0.1)	-0.019621	-0.190403	-0.151815	-3521152612319.554688	-0.469354
SVR(C=0.001, kernel='sigmoid')	-0.020827	-0.190505	-0.151952	-3504814566659.317383	-0.469531
DecisionTreeRegressor(criterion='absolute_error', max_depth=2, max_features=2, random_state=3000, splitter='random')	-0.014374	-0.189994	-0.151191	-3480337457527.190430	-0.468691



Обучение регрессионных моделей

Неудовлетворительные результаты





Анализ полученных данных



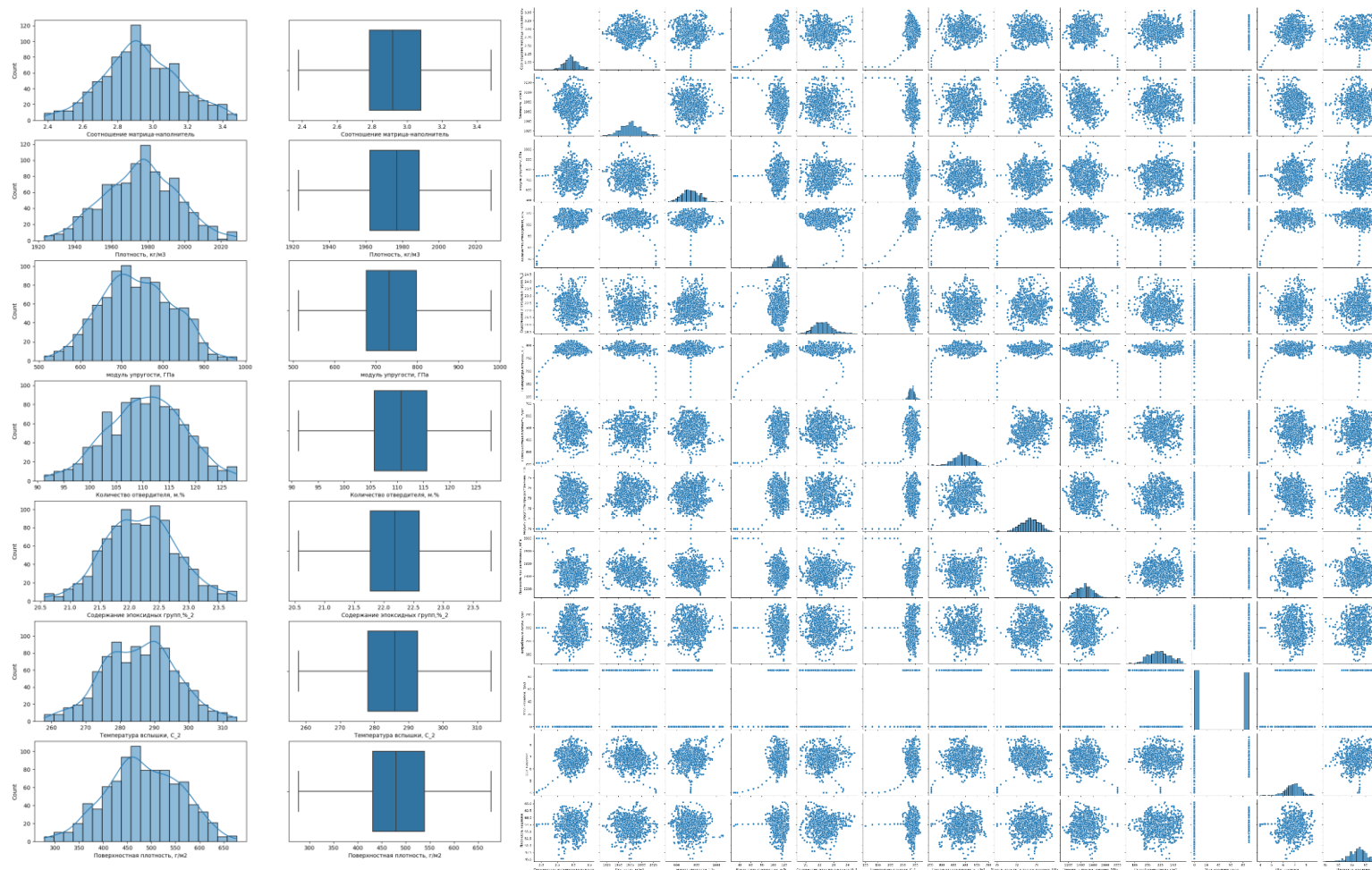
Дополнительная информация

- Кластеризация данных с заменой параметром главным компонентом по методу PCA





Удаление шумов из датасета



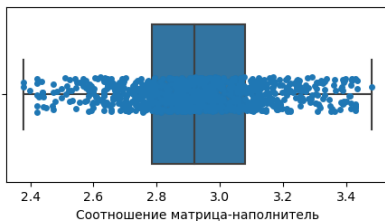
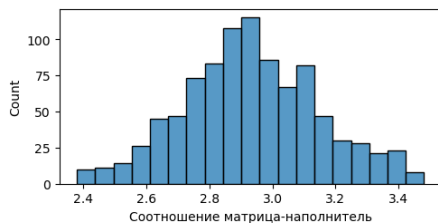
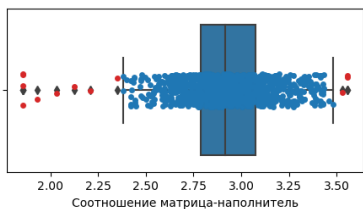
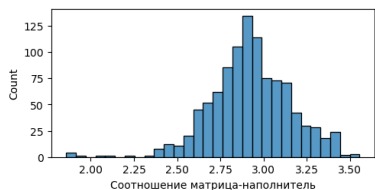
Дополнительная информация

- Удаление шумов медианным фильтром и фильтром Калмана

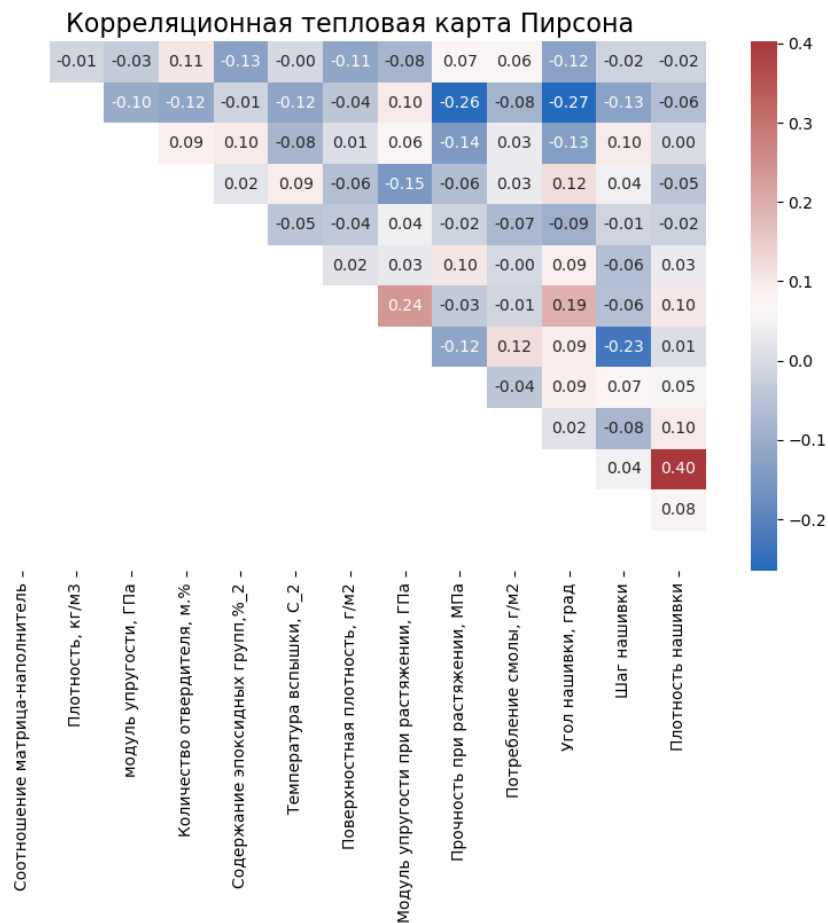




Удаление шумов из датасета



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 нашивки -



Дополнительная информация

- Удаление выбросов



Обучение регрессионных моделей

Модуль упругости при растяжении, ГПа и Прочность при растяжении, МПа

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.023790	-0.206570	-0.168089	-0.663950	-0.482383
LinearRegression	0.127503	-0.190622	-0.156021	-0.548095	-0.446032
Ridge	0.128356	-0.190538	-0.155920	-0.551555	-0.443519
Lasso	-0.023790	-0.206570	-0.168089	-0.663950	-0.482383
SVR	0.702189	-0.111048	-0.089952	-0.289343	-0.387441
KNeighborsRegressor	0.664991	-0.117629	-0.087594	-0.281993	-0.377483
DecisionTreeRegressor	0.131435	-0.189391	-0.120376	-0.363056	-0.687412
RandomForestRegressor	0.608477	-0.127583	-0.098812	-0.360453	-0.377803

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=10, solver='lsqr')	0.117797	-0.191737	-0.156796	-0.575752	-0.438060
Lasso(alpha=0.001)	0.117717	-0.191701	-0.156220	-0.564403	-0.449054
SVR(C=0.02, kernel='poly')	0.405342	-0.156877	-0.125619	-0.421175	-0.430320
KNeighborsRegressor(n_neighbors=3)	0.798118	-0.091141	-0.062416	-0.190646	-0.330939
DecisionTreeRegressor(criterion='absolute_error', max_depth=3, max_features=8, random_state=3000)	0.042181	-0.199153	-0.159513	-0.583590	-0.492044

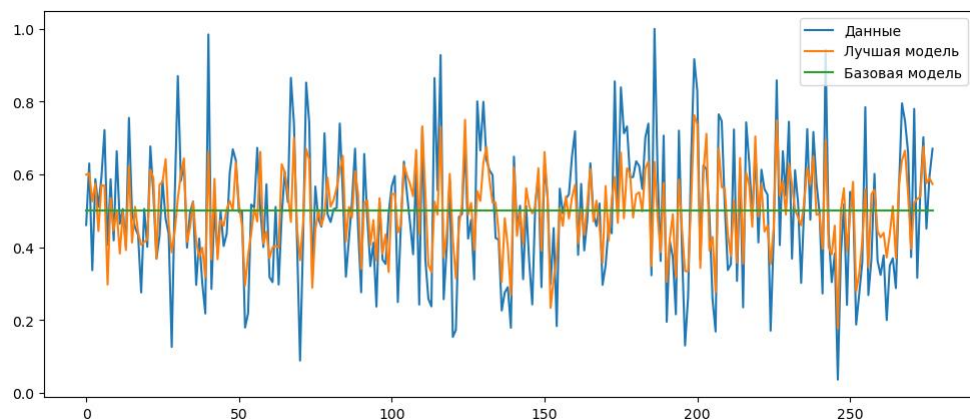
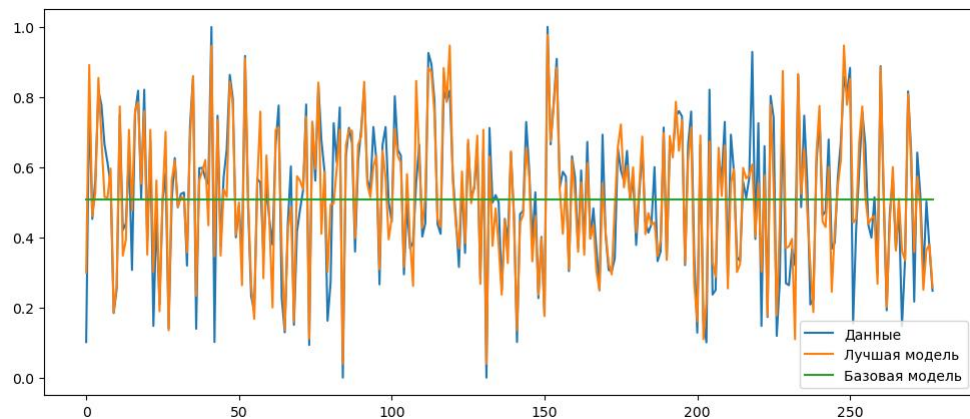
	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.014977	-0.184895	-0.150164	-3521621176373.428223	-0.439286
LinearRegression	0.084826	-0.175164	-0.141750	-2540445240983.733887	-0.433362
Ridge	0.086096	-0.175068	-0.141686	-2594856255449.057617	-0.430118
Lasso	-0.014977	-0.184895	-0.150164	-3521621176373.428223	-0.439286
SVR	0.647987	-0.108660	-0.085982	-1360307764636.201904	-0.330648
DecisionTreeRegressor	0.064370	-0.177365	-0.119034	-3047049950003.668945	-0.576177
GradientBoostingRegressor	0.396265	-0.142353	-0.112267	-2514647420319.952637	-0.389055

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=10, solver='lsqr')	0.078749	-0.175455	-0.141922	-2896413966050.510254	-0.408503
Lasso(alpha=0.1)	-0.016847	-0.184674	-0.150025	-3536279030688.050781	-0.425564
SVR(C=0.3)	0.512793	-0.127459	-0.100090	-1602416386036.175293	-0.348134
DecisionTreeRegressor(criterion='absolute_error', max_depth=3, max_features=5, random_state=3000)	0.080539	-0.175126	-0.138605	-2083710653417.593750	-0.421304
GradientBoostingRegressor(max_features=5, n_estimators=150, random_state=3000)	0.461828	-0.134253	-0.105258	-2106608649504.983643	-0.391175



Обучение регрессионных моделей

Удовлетворительные результаты

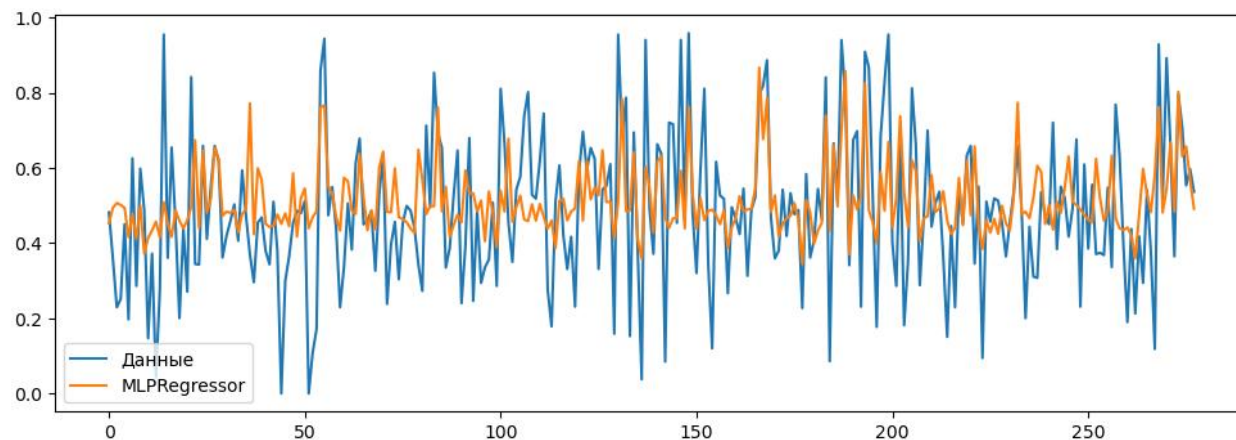




Обучение нейронных сетей

Разные типы нейросетей были обучены

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.014993	0.200563	0.157719	1.650086e+13	0.509286
MLPRegressor	0.218544	0.175983	0.135013	1.442920e+13	0.46255
Нейросеть переобученная	0.118761	0.186881	0.14111	5.037199e+12	0.520695
Нейросеть с ранней остановкой	0.208985	0.177056	0.135422	5.625733e+12	0.666826
Нейросеть dropout	-0.349995	0.231305	0.173923	3.329039e+12	0.791174





ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru