



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование конечных свойств новых материалов (композиционных материалов)

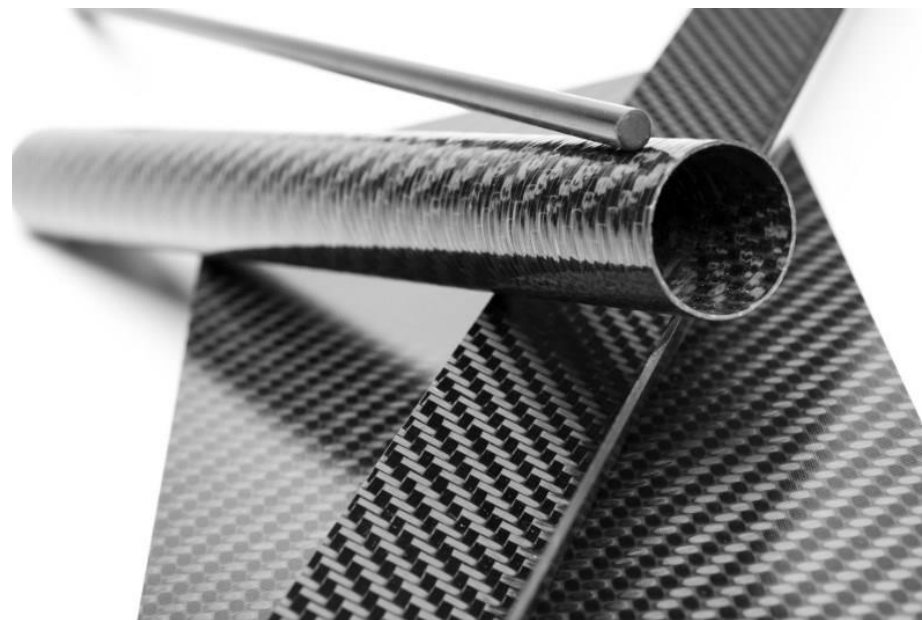
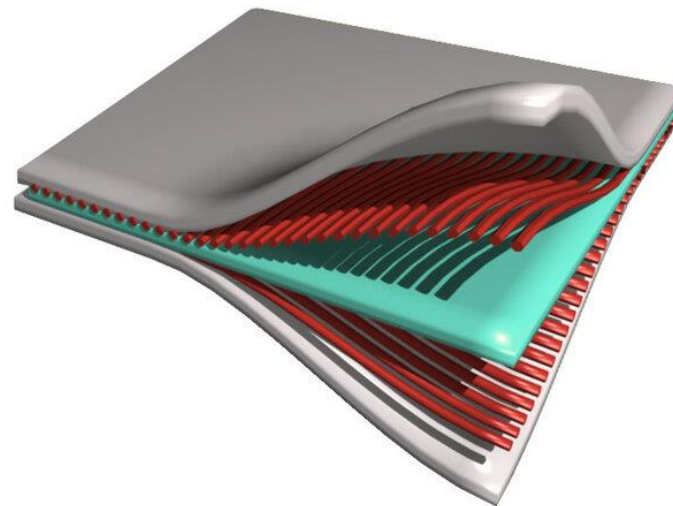
Семиврагов Сергей Александрович



Введение

Композиционные материалы

Композиционные материалы — это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними. Композиты обладают теми свойствами, которые не наблюдаются у компонентов по отдельности. При этом композиты являются монолитным материалом, т. е. компоненты материала неотделимы друг от друга без разрушения конструкции в целом. Яркий пример композита - железобетон.





Аналитическая часть

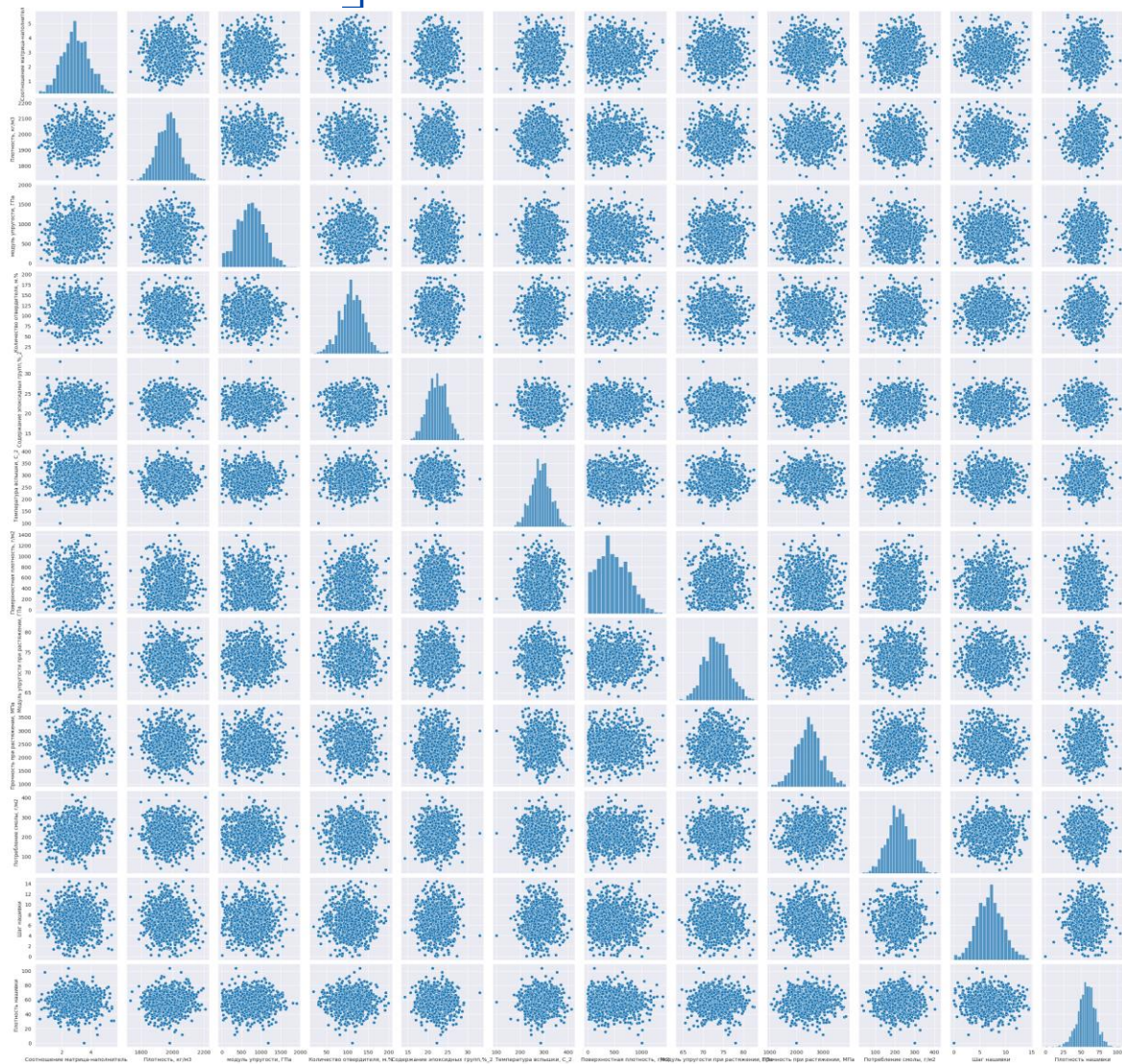
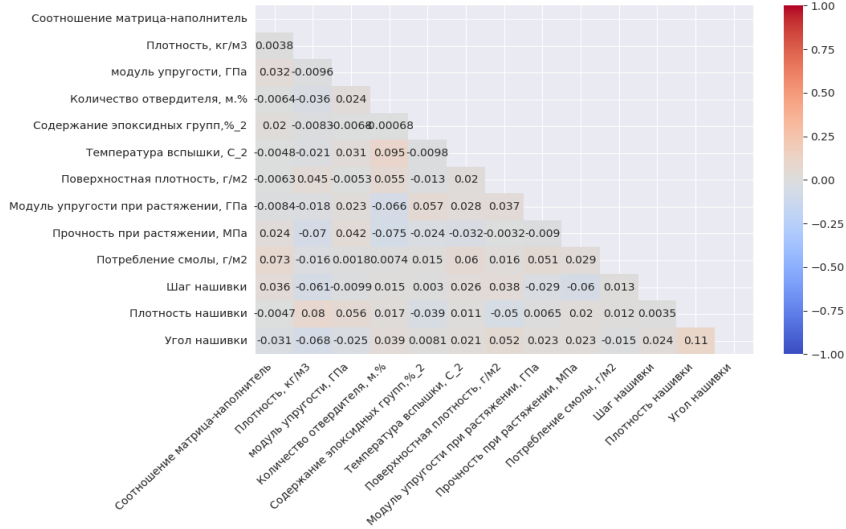
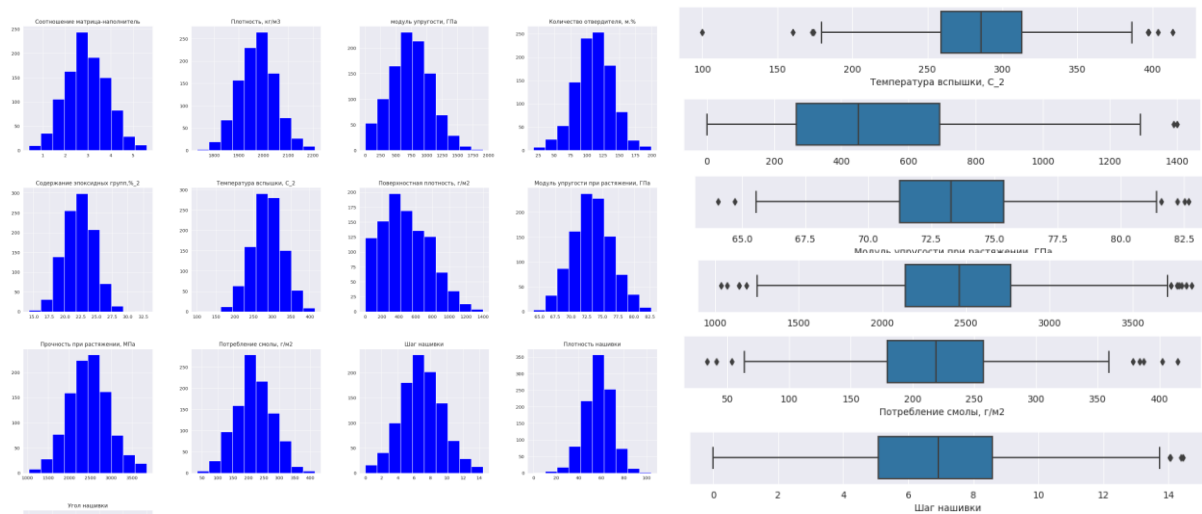
Используемые методы

№	Метод	Достоинства	Недостатки	Априорные предпосылки к работоспособности
1	К-ближайших соседей	Прост в реализации и понимании полученных результатов; имеет низкую чувствительность к выбросам; допускает настройку нескольких параметров; позволяет делать дополнительные допущения; универсален; находит лучшее решение из возможных; решает задачи небольшой размерности	Замедляется с ростом объема данных; не создаёт правил; не обобщает предыдущий опыт; основывается на всем массиве доступных исторических данных; невозможно сказать, на каком основании строятся ответы; сложно выбрать близость метрики; имеет высокую зависимость результатов классификации от выбранной метрики; полностью перебирает всю обучающую выборку при распознавании; имеет вычислительную трудоемкость	Прост в реализации и понимании
2	Стохастический и градиентный спуск	Эффективен; прост в реализации; имеет множество возможностей для настройки кода; способен обучаться на избыточно больших выборках	Требуется ряд гиперпараметров; чувствителен к масштабированию функций; может не сходиться или сходиться слишком медленно; функционал многоэкстремален; процесс может "застрять" в одном из локальных минимумов; возможно переобучение.	Простой, но очень эффективный подход к подгонке линейных регрессоров
3	Линейная регрессия	Быстр и прост в реализации; легко интерпретируем, имеет меньшую сложность по сравнению с другими алгоритмами	Моделирует только прямые линейные зависимости; требует прямую связь между зависимыми и независимыми переменными; выбросы оказывают огромное влияние, а границы линейны.	Прост с алгоритмической точки зрения.
4	Случайный лес (Random forest)	Не переобучается; не требует предобработки входных данных; эффективно обрабатывает пропущенные данные, данные с большим числом классов и признаков; имеет высокую точность предсказания и внутреннюю оценку обобщающей способности модели, а также высокую параллелизуемость и масштабируемость	Построение занимает много времени; сложно интерпретируемый; не обладает возможностью экстраполяции; может недообучаться; трудоёмко прогнозируемый; иногда работает хуже, чем линейные методы	Универсальный метод



Аналитическая часть

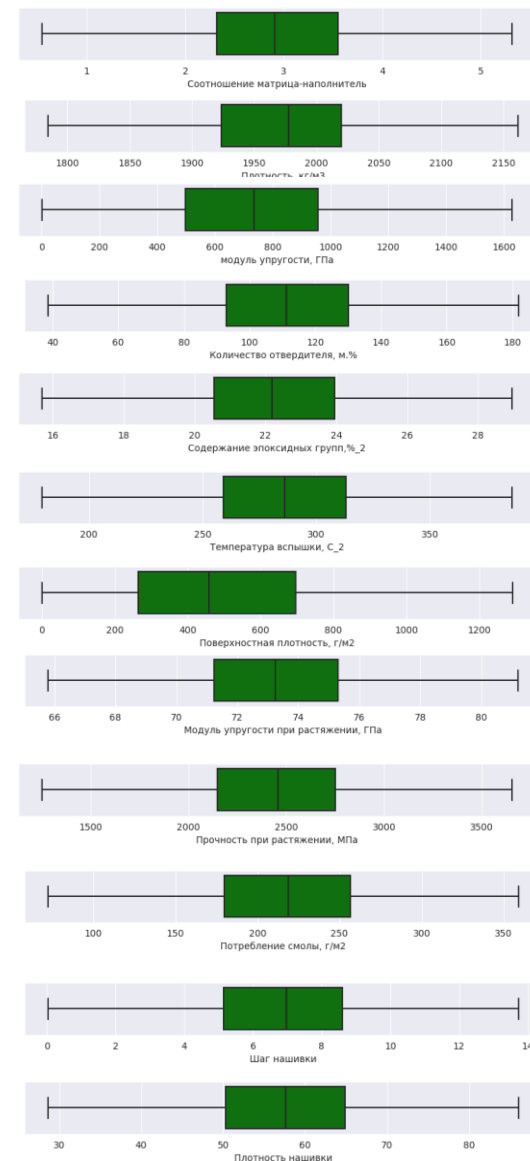
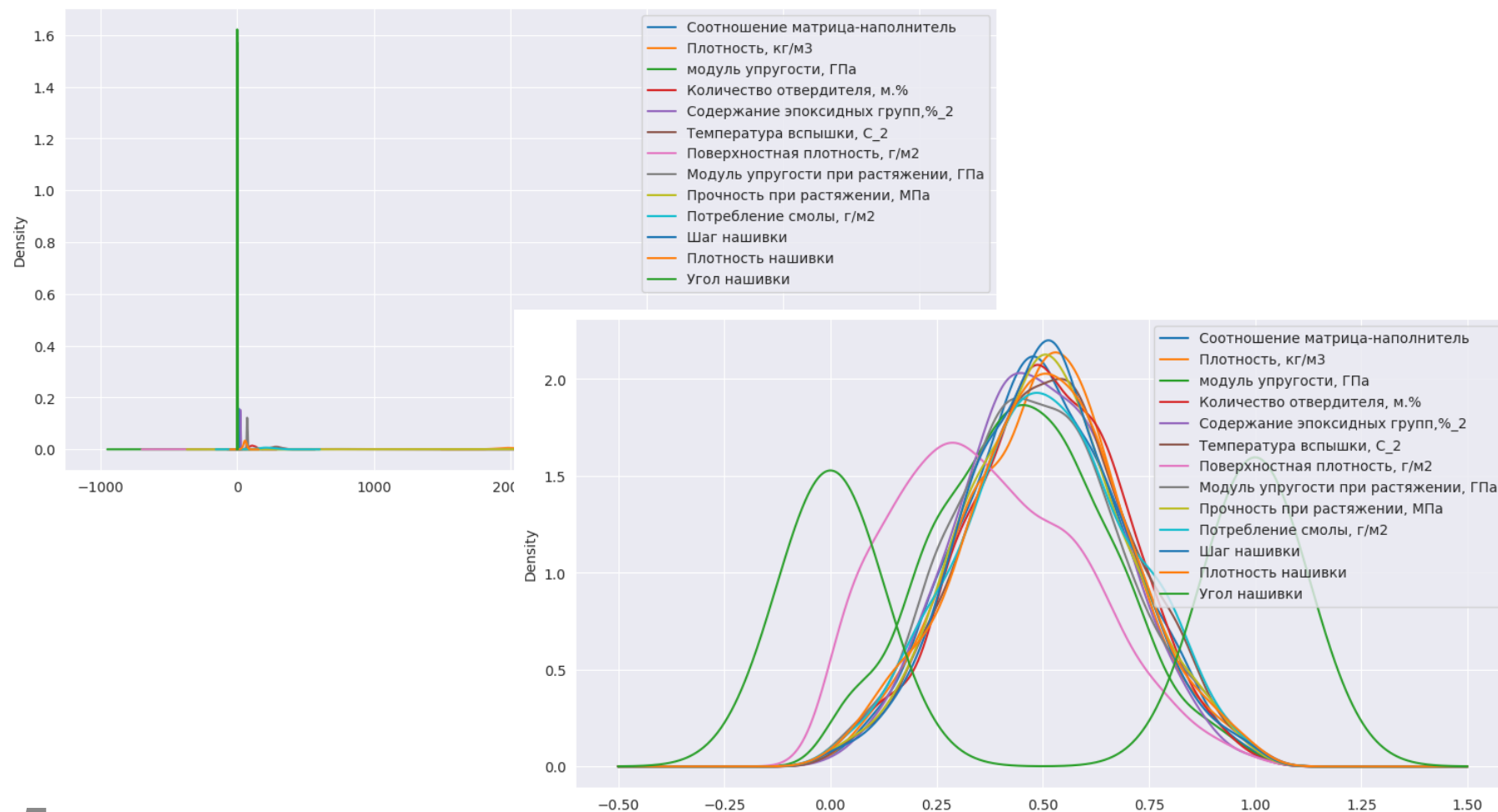
Разведочный анализ данных





Практическая часть

Предобработка данных

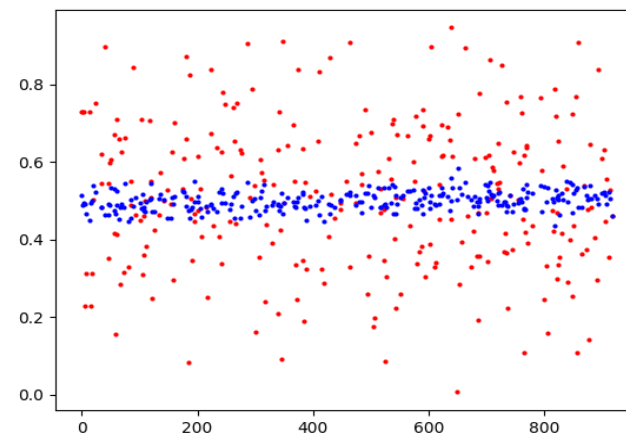
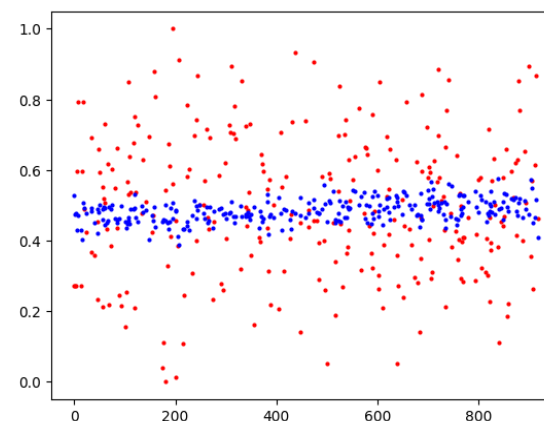
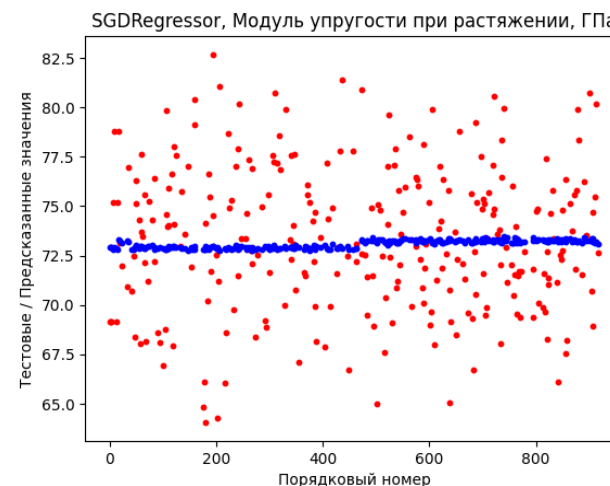
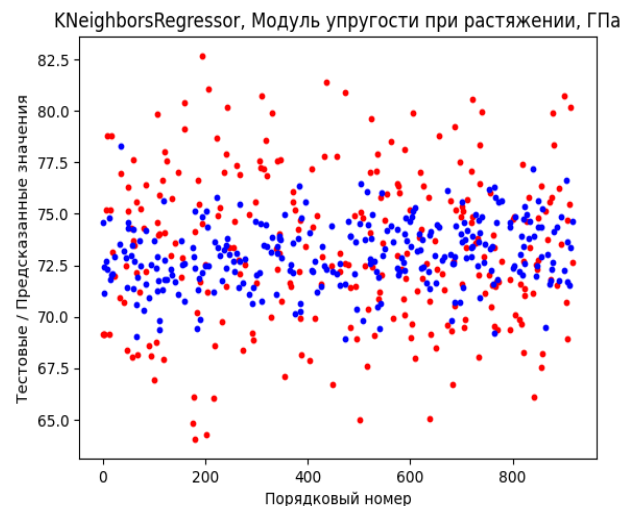




Практическая часть

Разработка и обучение модели

№	Модель	Искомое значение	MSE	MAE	R2 score
1	LinearRegression_pr	Прочность при растяжении	0.034523	0.149937	0.009
2	RandomForestRegressor_pr	Прочность при растяжении	0.035424	0.151399	-0.017
3	SGDRegressor_pr	Прочность при растяжении	0.035107	0.151721	-0.008
4	SGDRegressor_upr	Модуль упругости при растяжении	0.038705	0.159921	-0.011
5	LinearRegression_upr	Модуль упругости при растяжении	0.039564	0.161937	-0.034
6	RandomForestRegressor_upr	Модуль упругости при растяжении	0.040606	0.164424	-0.061
7	KNeighborsRegressor_pr	Прочность при растяжении	0.044102	0.165831	-0.266
8	KNeighborsRegressor_upr	Модуль упругости при растяжении	0.049561	0.181919	-0.295





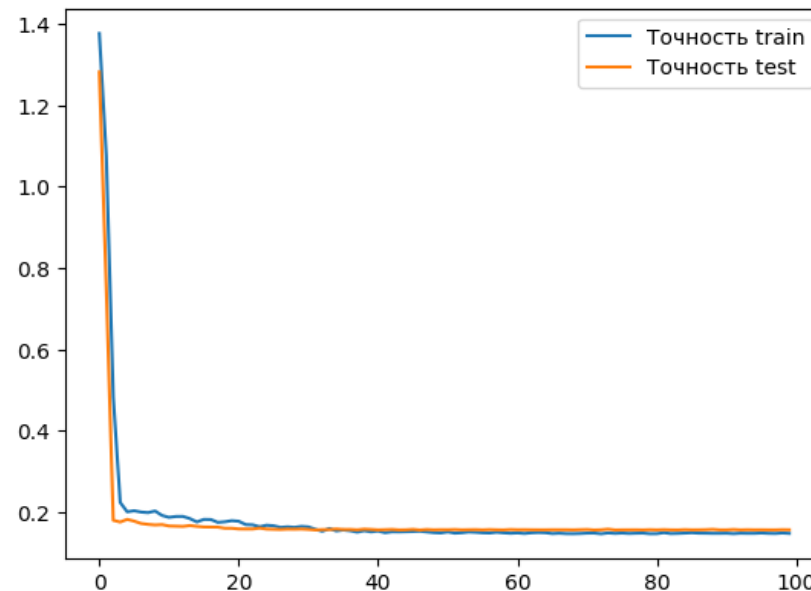
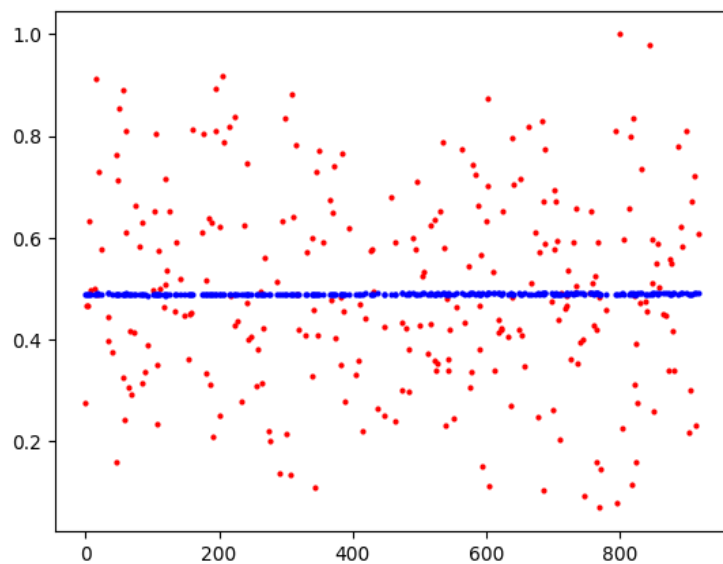
Практическая часть

Нейронная сеть, которая будет рекомендовать соотношение «матрица – наполнитель»

Model: "sequential_27"

Layer (type)	Output Shape	Param #
dense_109 (Dense)	(None, 32)	416
dropout_36 (Dropout)	(None, 32)	0
dense_110 (Dense)	(None, 64)	2112
dropout_37 (Dropout)	(None, 64)	0
dense_111 (Dense)	(None, 32)	2080
dense_112 (Dense)	(None, 16)	528
dense_113 (Dense)	(None, 1)	17

=====
Total params: 5,153
Trainable params: 5,153
Non-trainable params: 0



	MAE	MSE	R2
['Нейросеть 1 ',	0.15617427623257507,	0.037204866305288825,	-0.001247942991688289],
['Нейросеть 2 ',	0.1577630915770674,	0.037468339572840154,	-0.008338468865492388],
['Нейросеть 3 ',	0.15488497272542256,	0.037296911398320226,	-0.0037250372326391723],
['Нейросеть 4 ',	0.16182539498664764,	0.04181927417240977,	-0.1254297192993412],
['Нейросеть 5 ',	0.15555790582130452,	0.03715067493023215,	0.0002104415744100807]

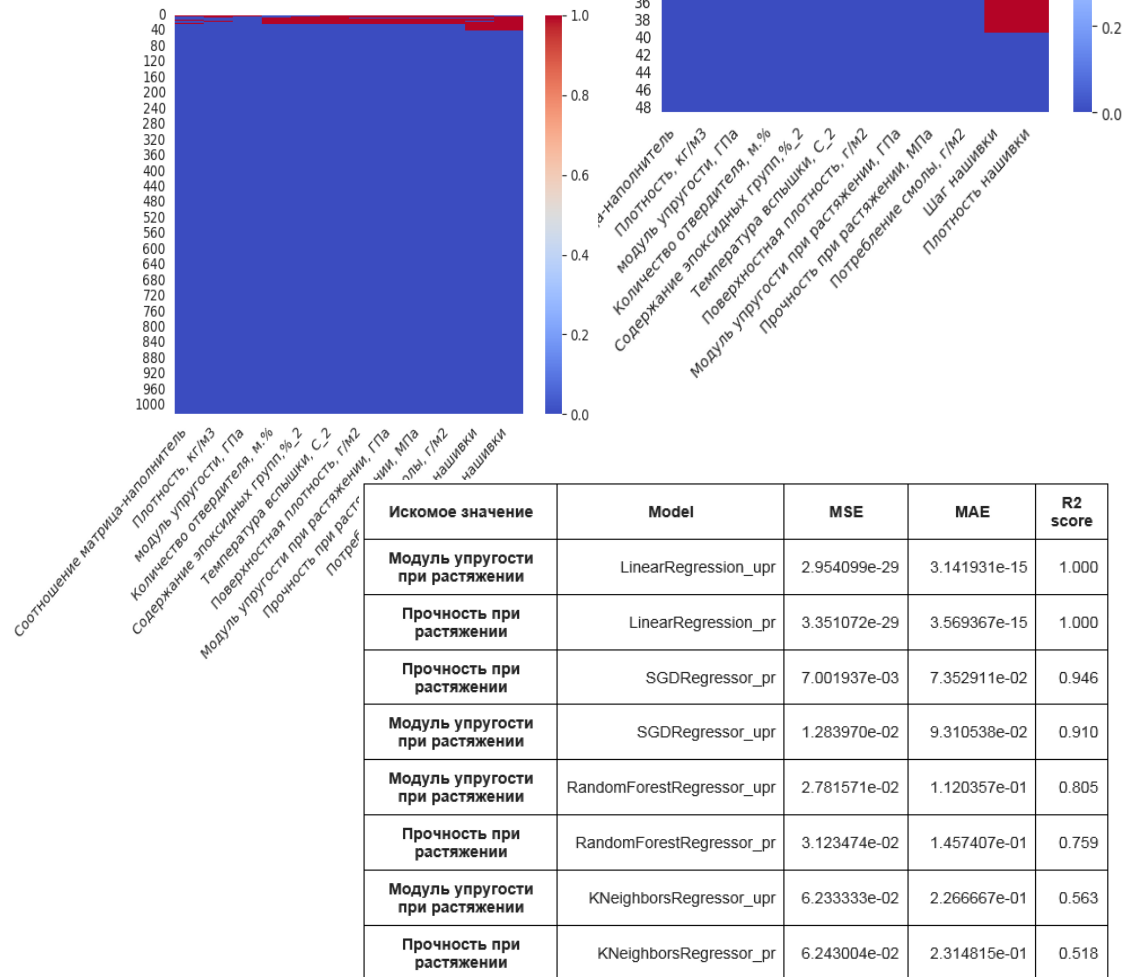


ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Практическая часть

Проверка датасета на вброс сгенерированных данных

Соотношение матрица- наполнитель	Плотность, кг/м3	модуль упругости, ГПа	Количество отвердителя, м. %	Содержание эпоксида, групп. % 2	Температура вспыхива, С	Поверхностная плотность, г/м2	Модуль упругости при растяжении, ГПа	Прочность при растяжении, МПа	Потребление смолы, г/м2	Шаг нашивки	Плотность нашивки	Угол нашивки	
0	1.857142857	2030	738.7368421	30	22,26785714	100	210	3000	220	4	57	0	
1	1.857142857	2030	738.7368421	50	23,75	284,6153846	210	70	3000	220	4	60	0
2	1.857142857	2030	738.7368421	49.9	33	284,6153846	210	70	3000	220	4	70	0
3	1.857142857	2030	738.7368421	129	21,25	300	210	70	3000	220	5	47	0
4	2.771331058	2030	753	111,86	22,26785714	284,6153846	210	70	3000	220	5	57	0
5	2.767918089	2000	748	111,86	22,26785714	284,6153846	210	70	3000	220	5	60	0
6	2.569620253	1910	807	111,86	22,26785714	284,6153846	210	70	3000	220	5	70	0
7	2.56147541	1900	535	111,86	22,26785714	284,6153846	380	75	1800	120	7	47	0
8	3.557017544	1930	889	129	21,25	300	380	75	1800	120	7	57	0
9	3.532338308	2100	1421	129	21,25	300	1010	78	2000	300	7	60	0
10	2.919677836	2160	933	129	21,25	300	1010	78	2000	300	7	70	0
11	2.877358491	1990	1628	129	21,25	300	1010	78	2000	300	9	47	0
12	1.598173516	1950	827	129	21,25	300	470	73,33333333	2455,555556	220	9	57	0
13	2.919677836	1980	568	129	21,25	300	470	73,33333333	2455,555556	220	9	60	0
14	4.029126214	1910	800	129	21,25	300	470	73,33333333	2455,555556	220	9	70	0
15	2.934782609	2030	302	129	21,25	300	210	70	3000	220	10	47	0
16	3.557017544	1880	313	129	21,25	300	210	70	3000	220	10	57	0
17	4.193548387	1950	506	129	21,25	300	380	75	1800	120	10	60	0
18	4.897959184	1890	540	129	21,25	300	380	75	1800	120	10	70	0
19	3.532338308	1980	1183	111,86	22,26785714	284,6153846	1010	78	2000	300	0	0	0
20	2.877358491	2000	205	111,86	22,26785714	284,6153846	1010	78	2000	300	4	47	90
21	1.598173516	1920	456	111,86	22,26785714	284,6153846	470	73,33333333	2455,555556	220	4	57	90
22	4.029126214	1880	622	111,86	22,26785714	284,6153846	470	73,33333333	2455,555556	220	4	60	90
23	2.587347643	1953.274926	1136.596135	137.6274196	22.34453557	234.716883	555.8934533	80.80322176	2387.342983	246.6131165	4	70	90
24	2.499917928	1942.595777	901.5199467	146.2522078	23.08175748	351.231874	864.7254838	76.17807508	3705.672523	226.2227604	5	47	90
25	2.046471464	2037.631811	707.570887	101.6172513	23.14639281	312.3072052	547.6012186	73.81706662	2624.026407	178.1985559	5	57	90
26	1.85647617	2018.220332	836.2943816	135.4016966	26.4355146	327.5103767	150.9614485	77.21076158	2473.187195	123.3445614	5	60	90
27	3.305335422	1917.907506	478.2862475	105.7869296	17.87409991	328.1545795	526.6921594	72.34570879	3059.032991	275.5758795	5	70	90
28	2.709534095	1892.071124	641.0525494	96.56329319	22.98929056	262.956722	804.5926208	74.51135922	2288.967377	126.8163389	7	47	90
29	2.282825314	2008.357592	593.9675252	149.3728324	21.66175068	330.498641	535.3714591	72.24492408	2704.445081	261.0770716	7	57	90
30	1.978140173	1973.629097	991.7240846	149.3721279	19.75057789	252.0581913	485.4537781	75.66570056	2448.943079	162.4936936	7	60	90
31	1.771436393	1872.49156	801.0338825	79.79434787	22.96630572	340.7368984	864.9291837	70.94759156	2796.785402	123.5362643	7	70	90
32	3.277086987	2010.047012	339.5304228	67.49899306	24.28069092	254.9490837	117.5352342	67.47870663	2462.605386	207.0185813	9	47	90
33	2.984362226	1912.315437	1183.091845	133.5490007	23.26379657	314.9961255	577.3890094	75.29045222	2303.770656	200.5802494	9	57	90
34	2.916149621	1879.969846	1003.270178	109.2395305	25.68275948	294.0485366	408.3542393	71.70085562	3086.546196	192.1911621	9	60	90
35	3.247617211	1813.2346	757.874479	81.57987084	23.42246524	279.0801575	575.0628571	69.34113288	3188.136358	252.8705688	9	70	90
36	2.423875673	1908.940601	530.2286864	38.26241428	24.07354923	325.138688	456.9080467	74.24435417	1890.505807	222.6994873	10	47	90
37	5.09899309	1977.339047	1572.096042	132.3430598	25.39700098	286.5564309	690.3648357	72.34163973	1386.578973	271.9013937	10	57	90
38	2.444176986	2085.495837	931.3106361	110.5648399	23.48713976	270.2867651	278.2300203	71.47906047	2740.229631	187.8613727	10	60	90
39	2.667696929	2078.894676	1542.168458	132.1474033	22.6501092	357.9728962	787.299217	76.47178847	2559.643047	163.902778	10	70	90
40	3.034399483	1968.401388	455.8710188	61.42129652	23.49072291	316.4145721	637.3768927	75.09037174	2848.490078	311.0523979	7.856166547	64.30196385	0
41	2.465204971	1936.099137	1056.554985	71.29405822	24.5233807	271.9756783	129.0771629	66.42079436	2868.586527	227.0225573	7.401542567	19.25053314	0
42	2.664388649	1996.159145	525.0577741	77.50688254	18.12610706	223.4086854	28.65810234	69.48977348	2220.587445	314.7766687	6.675780339	78.62329934	0
43	1.193529582	1965.929227	899.603701	102.9590686	19.56671624	225.8102229	871.0889548	73.45469452	2335.541792	91.04764633	7.32639832	38.17697532	0
44	2.914333275	2049.373404	382.2633585	81.35204737	16.39159473	233.2960627	561.9921308	69.81461516	2262.784366	303.0754524	8.32569922	46.04542763	0
45	4.315665782	1913.379677	822.9187355	143.5769371	24.2755881	274.9887944	260.859411	75.95732857	1639.912525	248.2443299	7.656210875	33.57102356	0
46	2.338424288	1963.35156	1155.160504	150.0158298	18.29942138	315.9041781	644.3631421	73.30398977	3407.713581	304.4232741	10.30294472	39.23427979	0
47	1.298167246	1984.511373	1405.786822	130.9427979	21.82932999	288.9520988	161.0077185	74.68091326	2526.814256	228.8677196	8.946891121	72.08459409	0
48	2.134464144	1986.349053	809.2938035	95.08902184	19.36495272	205.499761	196.3576427	76.34020725	2459.524526	289.9571416	3.746624977	57.9977218	0
49	4.147966432	1991.789789	1250.198275	116.8564622	21.57326495	320.7401725	755.5005552	70.35546403	1795.719359	189.8833073	9.094363677	44.80160057	0





ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Практическая часть

Разработка приложения Flask

Создание удаленного репозитория на GitHub

Sergey7OV / CompoZit

Рекомендации соотношения матрица-наполнитель для композиционных материалов

Введите параметры

<input type="text"/>	Плотность, кг/м3
<input type="text"/>	Модуль упругости, ГПа
<input type="text"/>	Количество отвердителя, м.%
<input type="text"/>	Содержание эпоксидных групп, % ₂
<input type="text"/>	Температура вспышки, C ₂
<input type="text"/>	Поверхностная плотность, г/м2
<input type="text"/>	Модуль упругости при растяжении, ГПа
<input type="text"/>	Прочность при растяжении, МПа
<input type="text"/>	Потребление смолы, г/м2
<input type="text"/>	Угол нашивки, град
<input type="text"/>	Шаг нашивки
<input type="text"/>	Плотность нашивки

Рассчитать Сбросить

The screenshot shows the GitHub repository page for 'Sergey7OV / CompoZit'. The repository is public and has 1 branch and 0 tags. The commit history shows a recent update to the README.md file. The repository description is 'Прогнозирование конечных свойств композиционных материалов' (Prediction of final properties of composite materials). The repository also includes a README.md file and a data directory.

File	Commit Message	Time
data	Data_CompoZit	9 hours ago
README.md	Update README.md	2 hours ago
VKR_clean_23.ipynb	Выявление и удаление сгенерированных данных	8 hours ago

README.md

Прогнозирование конечных свойств композиционных материалов

Выпускная квалификационная работа по курсу «Data Science» в ЦДО МГТУ им. Н. Э. Баумана по теме: "Прогнозирование конечных свойств новых материалов (композиционных материалов)". Описание:



Выводы

В процессе исследовательской работы были изучены теоретические основы и методы решения задачи регрессии с целью разработки модели для прогноза модуля упругости при растяжении, прочности при растяжении и соотношения «матрица-наполнитель».

Проведен разведочный анализ данных объединенного датасета, на основании которого сделан вывод, что распределение полученных данных близко к нормальному, корреляция между парами признаков близка к нулю.

Примененные методы решения задачи регрессии для объединенного и нормализованного датасета не привели к желаемому результату, коэффициент детерминации всех моделей меньше 0,5.

Исследовательская работа позволила сделать вывод, что часть предоставленных данных о начальных свойствах компонентов композиционных материалов была сгенерирована. Проведен анализ повторяемости данных в датасете, составлена тепловая карта и выявлена «зона» сгенерированных данных.

Сгенерированные данные в количестве 1000 строк были удалены и оставлены только данные полученные экспериментальным путем, что позволило на их основе провести повторное тестирование моделей. На основании модели линейной регрессии получен коэффициент детерминации равный 1.

Однако, очищенный от сгенерированных данных датасет содержит всего лишь 23 строки и его значения не подтверждены нормальному распределению. При этом выявлена высокая корреляция между значениями "Модуль упругости при растяжении, ГПа" и "Поверхностная плотность, г/м²"; значениями "Прочность при растяжении, МПа" и "Модуль упругости при растяжении, ГПа", а также значениями "Температура вспышки, С₂" и "Количество отвердителя, м.%".

В дальнейшем, для получения нормального распределения значений в датасете необходимы исследования для получения дополнительных экспериментальных значений начальных свойств компонентов композиционных материалов.



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru

Спасибо за внимание!