



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Выпускная квалификационная работа по курсу "Data Science"

Тема: Прогнозирование конечных свойств
новых материалов (композиционных материалов)

Слушатель: Несмиянов Сергей Витальевич



Постановка задачи и используемые методы выполнения

Методы

1 Линейная регрессия

2 Лассо-регрессия

3 Метод опорных
векторов

4 Дерево решений

5 Случайный лес

6 Градиентный бустинг

7 К-ближайших соседей

Целевые признаки

Модуль упругости при
растяжении

Прочность при
растяжении

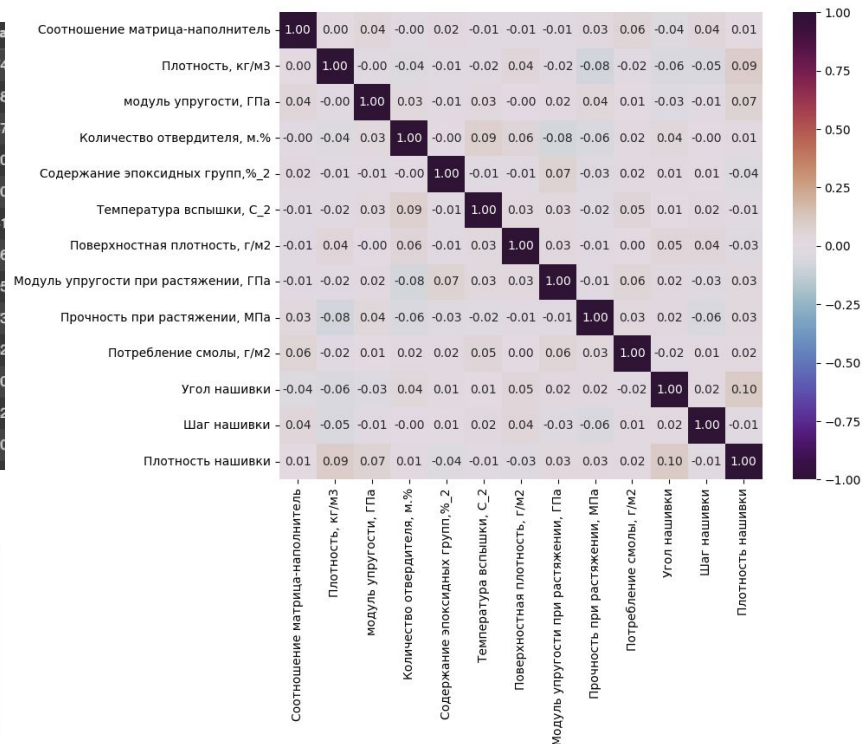
Матрица -
наполнитель

В ходе работы исследуется набор данных, состоящий из двух файлов-таблиц. Необходимо произвести объединение таблиц, разведочный анализ данных, описать методы исследования, осуществить предобработку данных, а также обучить алгоритм машинного обучения для определения значений двух заданных переменных и построить нейросеть для рекомендации соотношения третьей целевой переменной.

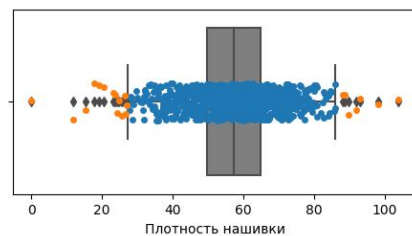
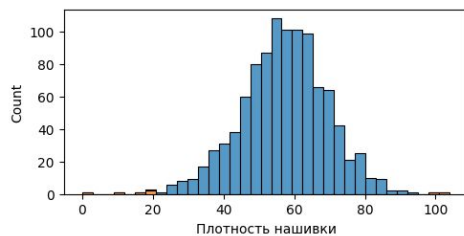


Разведочный анализ данных

	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	5.59174
Плотность, кг/м3	1023.0	1975.734888	73.729231	1731.764635	1924.155467	1977.621657	2021.374375	2207.77348
модуль упругости, ГПа	1023.0	739.923233	330.231581	2.436909	500.047452	739.664328	961.812526	1911.53647
Количество отвердителя, м.%	1023.0	110.570769	28.295911	17.740275	92.443497	110.564840	129.730366	198.95320
Содержание эпоксидных групп, %_2	1023.0	22.244390	2.406301	14.254985	20.608034	22.230744	23.961934	33.00000
Температура вспышки, С_2	1023.0	285.882151	40.943260	100.000000	259.066528	285.896812	313.002106	413.27347
Поверхностная плотность, г/м2	1023.0	482.731833	281.314690	0.603740	266.816645	451.864365	693.225017	1399.54236
Модуль упругости при растяжении, ГПа	1023.0	73.328571	3.118983	64.054061	71.245018	73.268805	75.356612	82.68205
Прочность при растяжении, МПа	1023.0	2466.922843	485.628006	1036.856605	2135.850448	2459.524526	2767.193119	3848.43673
Потребление смолы, г/м2	1023.0	218.423144	59.735931	33.803026	179.627520	219.198882	257.481724	414.59062
Угол нашивки	1023.0	0.491691	0.500175	0.000000	0.000000	0.000000	1.000000	1.00000
Шаг нашивки	1023.0	6.899222	2.563467	0.000000	5.080033	6.916144	8.586293	14.44052
Плотность нашивки	1023.0	57.153929	12.350969	0.000000	49.799212	57.341920	64.944961	103.98890



Описательная статистика объединенного набора данных



Пример поиска выбросов

Матрица корреляции признаков



Предобработка (препроцессинг)

```
# Описательная статистика входных данных до предобработки
```

```
show_statistics(xl_train_raw)
```

	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	Шаг нашивки	Плотность нашивки
min	0.547391	1784.482245	2.436909	33.624187	15.695894	173.973907	1.668002	41.048278	0.000000	0.037639	20.571633
max	5.591742	2192.738783	1649.415706	192.851702	28.907470	403.652861	1288.691844	386.903431	1.000000	14.033215	92.963492
mean	2.943860	1972.286516	738.627618	112.119243	22.179055	286.449560	481.805877	216.838475	0.495714	6.880379	57.403269
std	0.902194	73.148332	326.130594	28.056458	2.335087	40.645101	278.253589	58.108052	0.500339	2.590968	12.036623

```
# Описательная статистика входных данных после предобработки
```

```
show_statistics(pd.DataFrame(xl_train, columns=(xl_continuous + x_categorical)))
```

	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	Шаг нашивки	Плотность нашивки	Угол нашивки
min	-2.658166	-2.569280	-2.258964	-2.799754	-2.778397	-2.769241	-1.726774	-3.027393	-2.642886	-3.062152	0.000000
max	2.937033	3.015925	2.794707	2.879558	2.883502	2.885639	2.901896	2.928795	2.762655	2.956448	1.000000
mean	0.000000	0.000000	-0.000000	-0.000000	-0.000000	0.000000	0.000000	-0.000000	0.000000	-0.000000	0.495714
std	1.000715	1.000715	1.000715	1.000715	1.000715	1.000715	1.000715	1.000715	1.000715	1.000715	0.500339

Пример описательной статистики входных признаков до и после предобработки

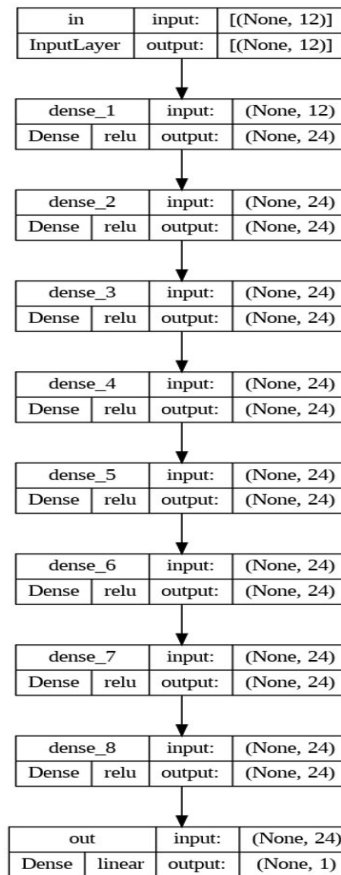


Подбор моделей и построение нейронной сети

Ridge(alpha=80, positive=True, solver='lbfgs')	-0.016604	-3.121555	-2.494491	-0.034068	-7.581330
Lasso(alpha=0.05)	-0.018084	-3.114388	-2.500839	-0.034157	-7.965382
SVR(C=0.01, kernel='llinear')	-0.017814	-3.123659	-2.500515	-0.034147	-8.061850
KNeighborsRegressor(n_neighbors=29)	-0.036593	-3.147992	-2.512539	-0.034342	-8.157406
DecisionTreeRegressor(max_depth=2, max_features=2, random_state=42)	-0.018267	-3.125442	-5.490188	-0.034017	-8.154902
RandomForestRegressor(bootstrap=False, criterion='absolute_error', max_depth=5, max_features=1, n_estimators=50, random_state=42)	-0.017821	-3.124964	-2.497013	-0.034103	-8.146335

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=710, solver='sparse_cg')	-0.010171	-490.370725	-389.300830	-0.170584	-1291.112095
Lasso(alpha=20)	-0.010149	-490.357498	-389.214674	-0.170515	-1293.747766
SVR(C=0.02, kernel='llinear')	-0.021227	-493.150121	-390.526504	-0.170386	-1279.607766
DecisionTreeRegressor(max_depth=1, max_features=6, random_state=42, splitter='random')	-0.019887	-492.777398	-390.140019	-0.171163	-1281.728096
GradientBoostingRegressor(max_depth=2, max_features=1, random_state=42)	0.019853	-485.855328	-386.455908	-0.168528	-1276.277567

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.000744	-0.924041	-0.739327	-0.340221	-2.554458
Нейросеть переобученная	-0.702315	-1.205174	-0.971796	-0.396919	-3.332204
Нейросеть с ранней остановкой	-0.357113	-1.076064	-0.853269	-0.330199	-2.936997
Нейросеть dropout	-0.377819	-1.084241	-0.847411	-0.342221	-3.708873





Оценка точности работы

	R2	RMSE	MAE	MAPE	max_error
Модуль упругости, тренировочный	0.029799	-3.079407	-2.464290	-0.033661	-8.937810
Модуль упругости, тестовый	-0.021771	-3.104193	-2.512508	-0.034472	-8.816359

	R2	RMSE	MAE	MAPE	max_error
Прочность при растяжении, тренировочный	0.056663	-479.002875	-379.538641	-0.166241	-1369.735087
Прочность при растяжении, тестовый	-0.006114	-465.833555	-364.992369	-0.162767	-1399.342924

	R2	RMSE	MAE	MAPE	max_error
Соотношение матрица-наполнитель, тренировочный	-0.261320	-1.012517	-0.773151	-0.266820	-3.698861
Соотношение матрица-наполнитель, тестовый	-0.357113	-1.076064	-0.853269	-0.330199	-2.936997



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru