



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Мезенцева Екатерина Михайловна



Постановка задачи

1

Цель исследования: решение актуальной производственной задачи по прогнозированию свойств получаемых композиционных материалов.

2

Объект исследования: композиционные материалы, которые означают искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними.

3

Предмет исследования: прогнозные данные трех свойств композитов: соотношение матрица-наполнитель, модуль упругости при растяжении, прочность при растяжении.

Актуальность

Возможность использования цифрового подхода к процессу создания композиционных материалов для снижения количества реально проводимых испытаний.



Задачи:

1. Изучить теоретические основы и методы решения поставленной задачи;
2. Провести разведочный анализ предложенных данных. Необходимо нарисовать гистограммы распределения каждой из переменной, диаграммы ящика с усами, попарные графики рассеяния точек. Необходимо также для каждой колонке получить среднее, медианное значение, провести анализ и исключение выбросов, проверить наличие пропусков;
3. Провести предобработку данных (удаление шумов, нормализация и т.д.);
4. Обучить 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. При построении моделей провести поиск гиперпараметров модели с помощью поиска по сетке с перекрестной проверкой, количество блоков равно 10;
5. Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель;
6. Разработать приложение с графическим интерфейсом или интерфейсом командной строки, которое будет выдавать прогноз, полученный в задании 4 или 5 (один или два прогноза, на выбор учащегося);
7. Оценить точность модели на тренировочном и тестовом датасете;
8. Создать репозиторий в GitHub / GitLab и разместить там код исследования. Оформить файл README.



Преобразования и характеристики датасета

Датасеты объединены, тип объединения INNER.

Пропуски отсутствуют. Элементы массива соответствуют типу float64.

За исключением признака «Угол нашивки», все признаки в датасете являются непрерывными, количественными.

Признак «Угол нашивки» принимает только два значения и потому будет рассматриваться как категориальный признак.

Для каждой колонки получены среднее, стандартное отклонение, минимальное, максимальное, первый квартиль, медианное значение, третий квартиль.

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	1023.0	2.930	0.913	0.389	2.318	2.907	3.553	5.592
Плотность, кг/м3	1023.0	1975.735	73.729	1731.765	1924.155	1977.622	2021.374	2207.773
модуль упругости, ГПа	1023.0	739.923	330.232	2.437	500.047	739.664	961.813	1911.536
Количество отвердителя, м.%	1023.0	110.571	28.296	17.740	92.443	110.565	129.730	198.953
Содержание эпоксидных групп,%_2	1023.0	22.244	2.406	14.255	20.608	22.231	23.962	33.000
Температура вспышки, C_2	1023.0	285.882	40.943	100.000	259.067	285.897	313.002	413.273
Поверхностная плотность, г/м2	1023.0	482.732	281.315	0.604	266.817	451.864	693.225	1399.542
Модуль упругости при растяжении, ГПа	1023.0	73.329	3.119	64.054	71.245	73.269	75.357	82.682
Прочность при растяжении, МПа	1023.0	2466.923	485.628	1036.857	2135.850	2459.525	2767.193	3848.437
Потребление смолы, г/м2	1023.0	218.423	59.736	33.803	179.628	219.199	257.482	414.591
Угол нашивки, град	1023.0	0.492	0.500	0.000	0.000	0.000	1.000	1.000
Шаг нашивки	1023.0	6.899	2.563	0.000	5.080	6.916	8.586	14.441
Плотность нашивки	1023.0	57.154	12.351	0.000	49.799	57.342	64.945	103.989

Рисунок 1 – среднее, стандартное отклонение, минимальное, максимальное, первый квартиль, медианное значение, третий квартиль



Разведочный анализ данных

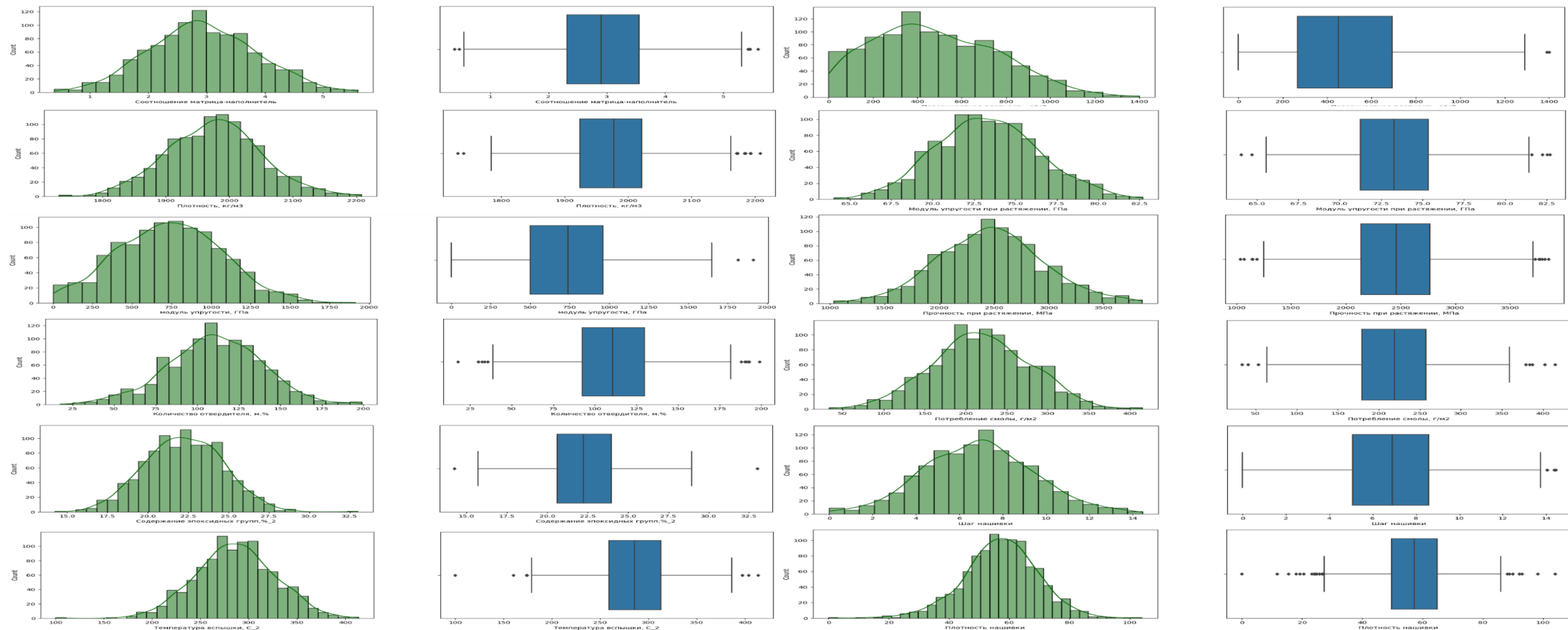


Рисунок 3 – гистограммы распределения и ящики с усами до нормализации данных, исключая угол нашивки как категориальный признак



Разведочный анализ данных

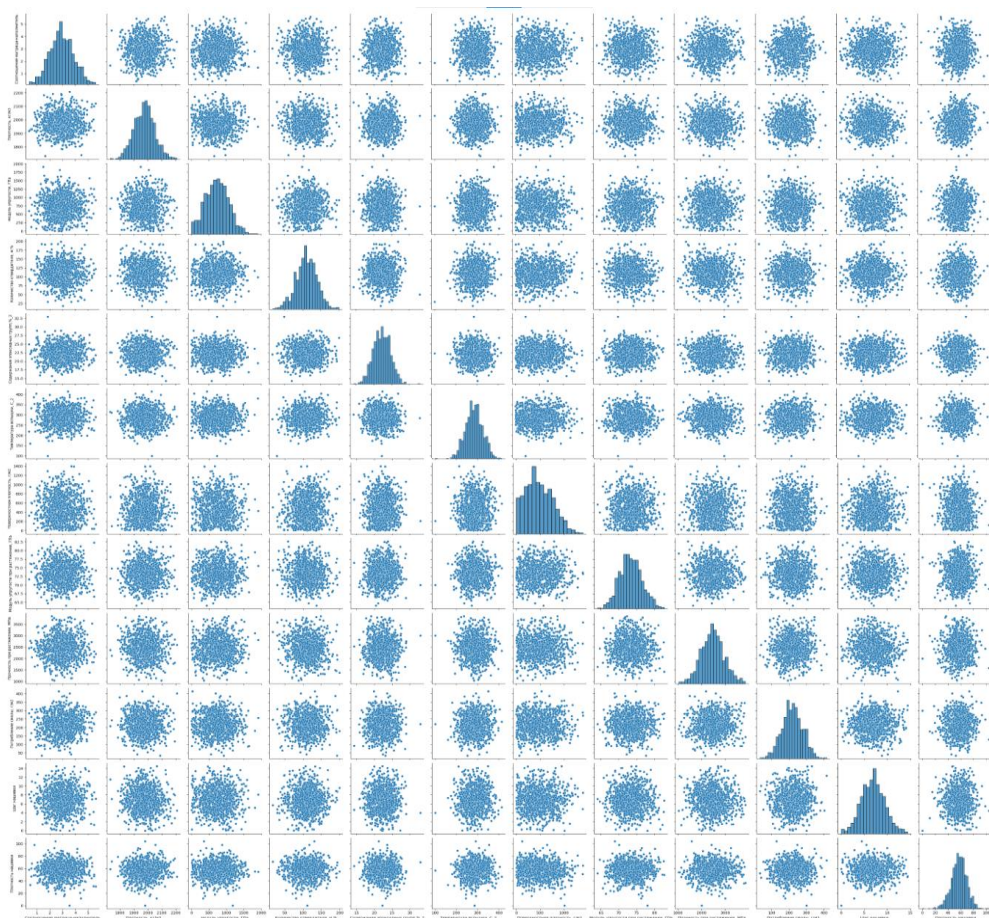


Рисунок 5 – попарные графики рассеяния точек до нормализации

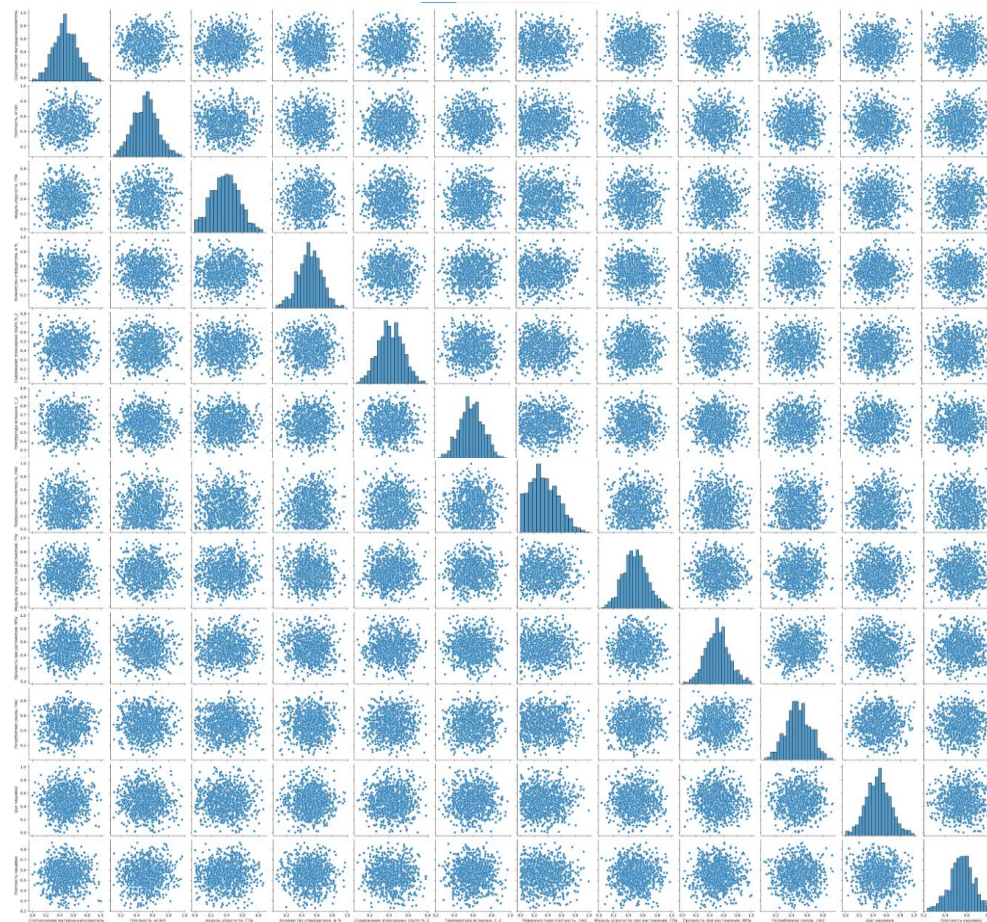


Рисунок 5 – попарные графики рассеяния точек после нормализации



Разведочный анализ данных

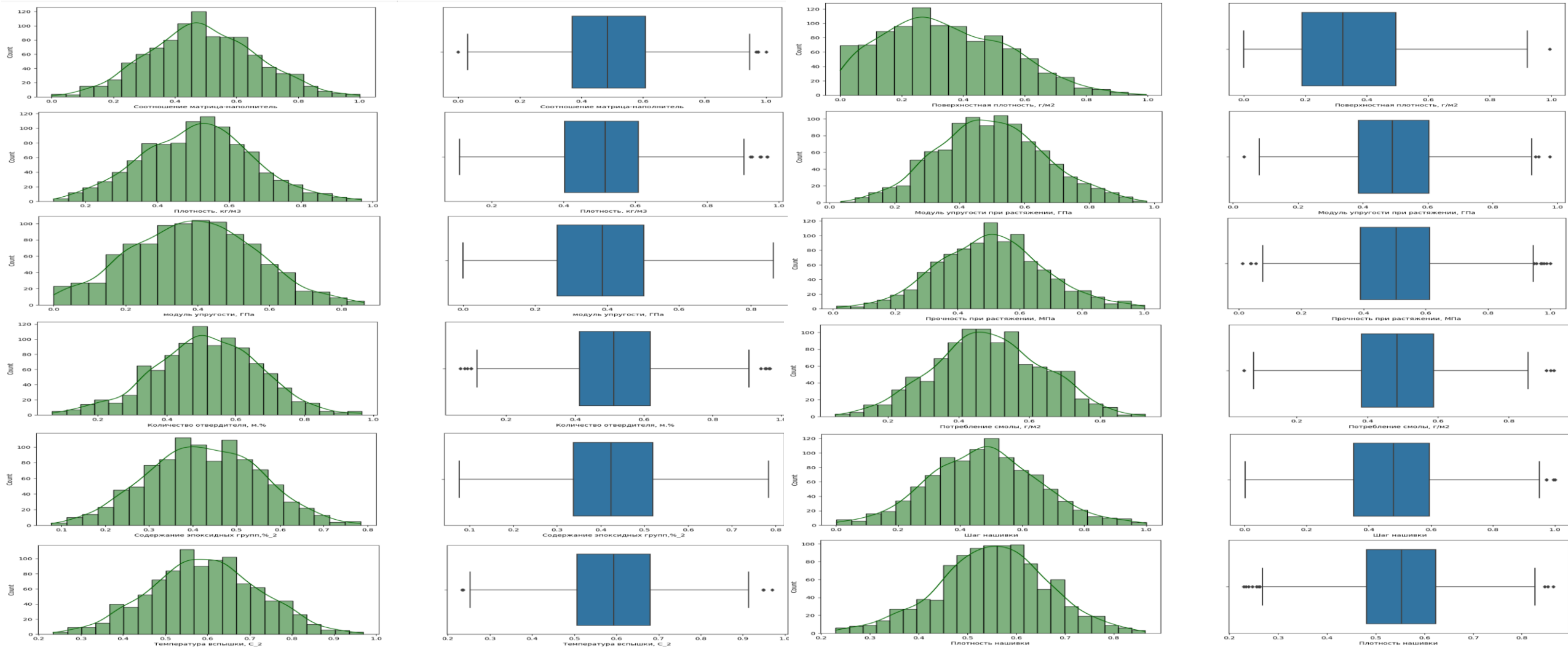


Рисунок 6 – гистограммы распределения и ящики с усами после нормализации данных и удаления выбросов, исключая угол нашивки как категориальный признак



Разведочный анализ данных

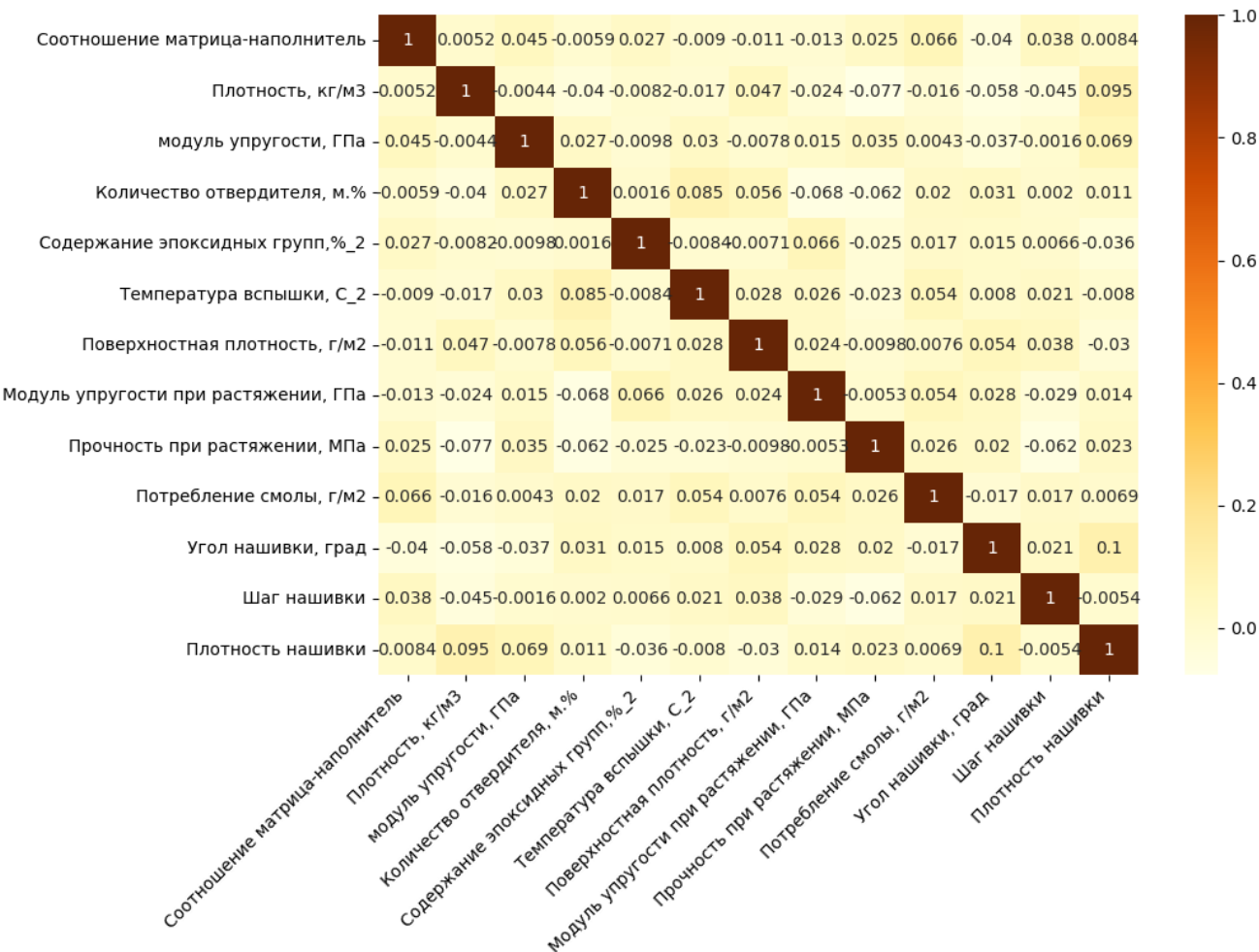


Рисунок 7 – тепловая карта коэффициентов корреляции

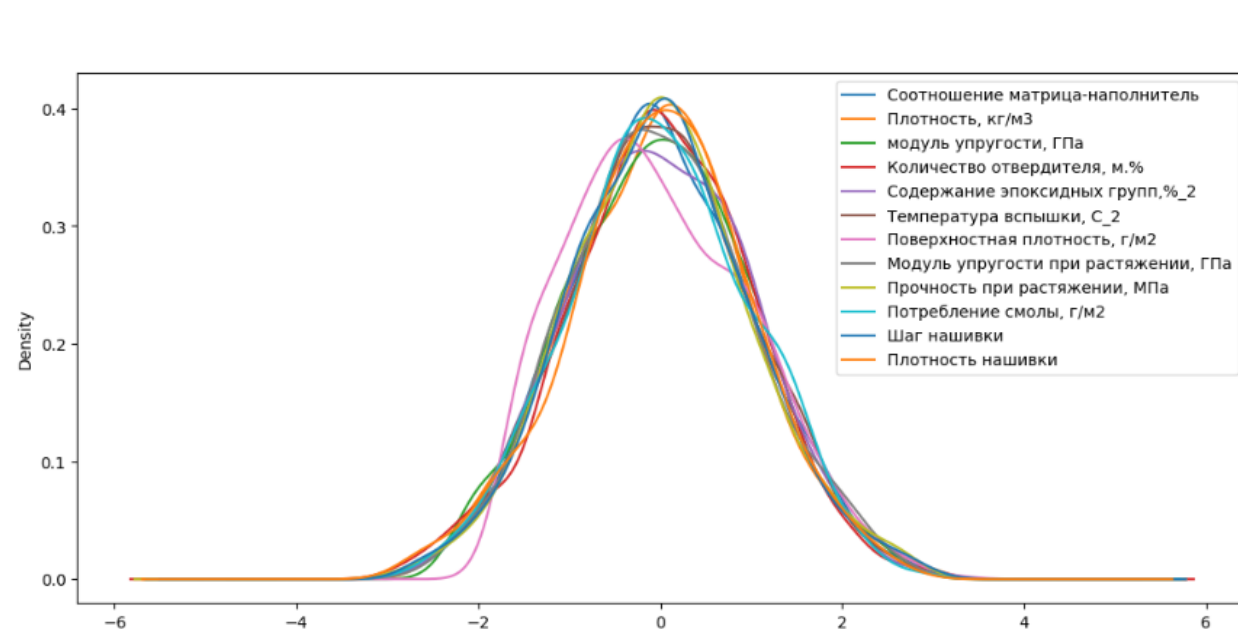


Рисунок 8 – визуализированные данные после нормализации



Разработка и обучение модели

Прогнозирование модуля упругости при растяжении

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.025801	-3.013795	-2.408483	-0.032937	-7.702150
LinearRegression	-0.040280	-3.033795	-2.419878	-0.033087	-7.785673
Ridge	-0.040201	-3.033684	-2.419806	-0.033087	-7.785518
Lasso	-0.025801	-3.013795	-2.408483	-0.032937	-7.702150
SVR	-0.064043	-3.067918	-2.446795	-0.033437	-7.896264
KNeighborsRegressor	-0.218550	-3.279741	-2.632233	-0.035947	-8.352322
DecisionTreeRegressor	-1.228511	-4.423601	-3.583321	-0.048992	-11.451281
RandomForestRegressor	-0.087091	-3.100472	-2.472667	-0.033824	-8.005775
GradientBoostingRegressor	-0.155287	-3.197943	-2.524645	-0.034507	-8.531914

Рисунок 9 – метрики работы выбранных моделей с гиперпараметрами по умолчанию

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=410, positive=True, solver='lbfgs')	-0.028041	-3.016480	-2.402560	-0.032854	-7.739427
Lasso(alpha=0.1)	-0.028228	-3.017201	-2.405317	-0.032891	-7.738947
SVR(C=0.05, kernel='poly')	-0.021899	-3.007710	-2.397686	-0.032754	-7.738811
KNeighborsRegressor(n_neighbors=53)	-0.037535	-3.030675	-2.416598	-0.033056	-7.720174
DecisionTreeRegressor(max_depth=2, max_features=5, random_state=0, splitter='random')	-0.011775	-2.991498	-2.387448	-0.032646	-7.755630
RandomForestRegressor(bootstrap=False, criterion='absolute_error', max_depth=4, max_features=1, n_estimators=50, random_state=0)	-0.035630	-3.028397	-2.411007	-0.032948	-7.862964
GradientBoostingRegressor(max_depth=1, max_features=1, n_estimators=50, random_state=0)	-0.046504	-3.044165	-2.420581	-0.033099	-7.861759

Рисунок 10 – метрики работы выбранных моделей с настроенными гиперпараметрами



Разработка и обучение модели

Прогнозирование прочности при растяжении

	R2	RMSE	MAE	MAPE	max_error
DummyRegressor	-0.026342	-484.807293	-380.185919	-0.167277	-1253.431489
LinearRegression	-0.033167	-486.474961	-385.238945	-0.168730	-1258.639354
Ridge	-0.033108	-486.460907	-385.220677	-0.168723	-1258.607692
Lasso	-0.032865	-486.404296	-384.988615	-0.168664	-1258.645065
SVR	-0.026718	-484.947303	-380.218028	-0.166329	-1262.072451
DecisionTreeRegressor	-1.069986	-682.369273	-545.666323	-0.234120	-1773.372125
RandomForestRegressor	-0.078820	-496.382732	-399.731376	-0.174450	-1243.269911
GradientBoostingRegressor	-0.148984	-511.332319	-409.803025	-0.178062	-1291.511611

Рисунок 11 – метрики работы выбранных моделей с гиперпараметрами по умолчанию

	R2	RMSE	MAE	MAPE	max_error
Ridge(alpha=990, positive=True, solver='lbfgs')	-0.023267	-484.113037	-380.018387	-0.167120	-1257.022616
Lasso(alpha=50)	-0.026342	-484.807293	-380.185919	-0.167277	-1253.431489
SVR(C=0.001, kernel='sigmoid')	-0.026311	-484.853433	-379.936358	-0.166191	-1261.977540
DecisionTreeRegressor(criterion='absolute_error', max_depth=1, max_features=5, random_state=0)	-0.026397	-484.768578	-378.716225	-0.166234	-1255.193290
RandomForestRegressor(bootstrap=False, criterion='absolute_error', max_depth=2, max_features=1, n_estimators=50, random_state=0)	-0.020263	-483.477346	-380.126241	-0.166222	-1248.234411
GradientBoostingRegressor(max_depth=1, max_features=1, n_estimators=50, random_state=0)	-0.032345	-486.308546	-382.902477	-0.168049	-1244.425207

Рисунок 12 – метрики работы выбранных моделей с настроенными гиперпараметрами



Визуализация работы нейросетей

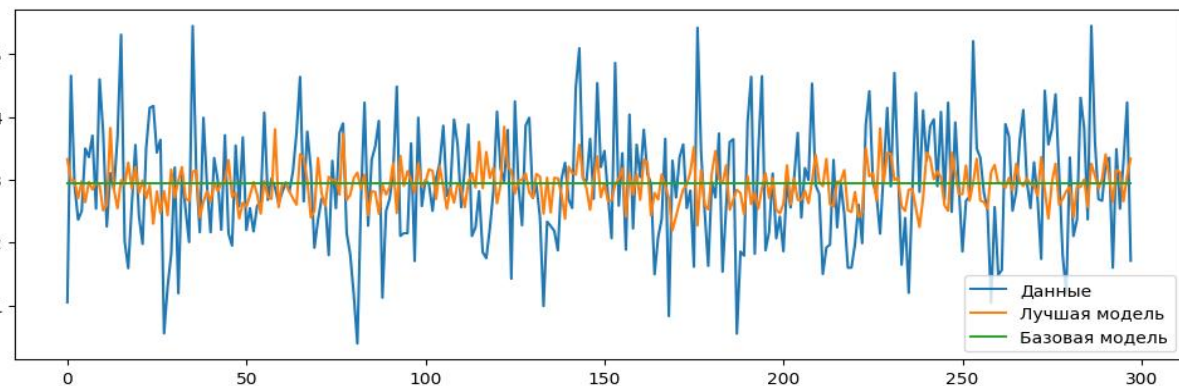


Рисунок 13 – визуализация работы модели MLPRegressor

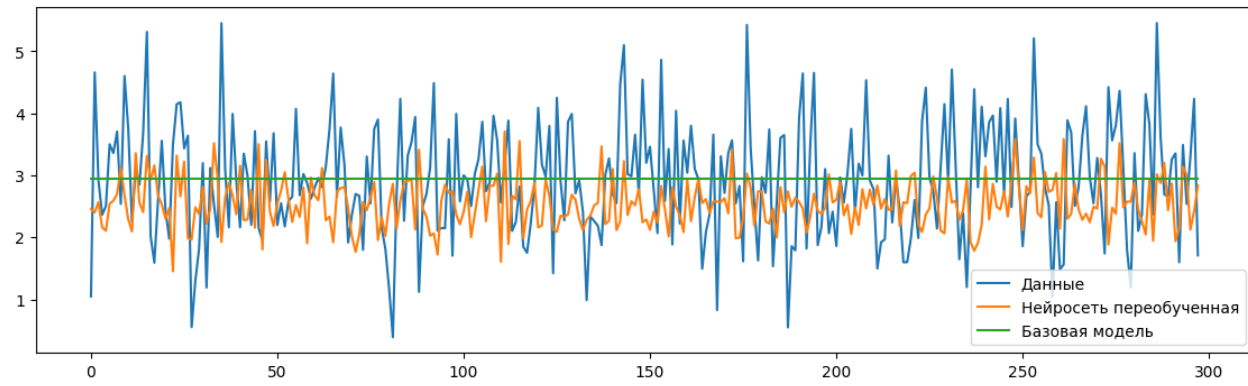


Рисунок 14 – визуализация работы переобученной модели

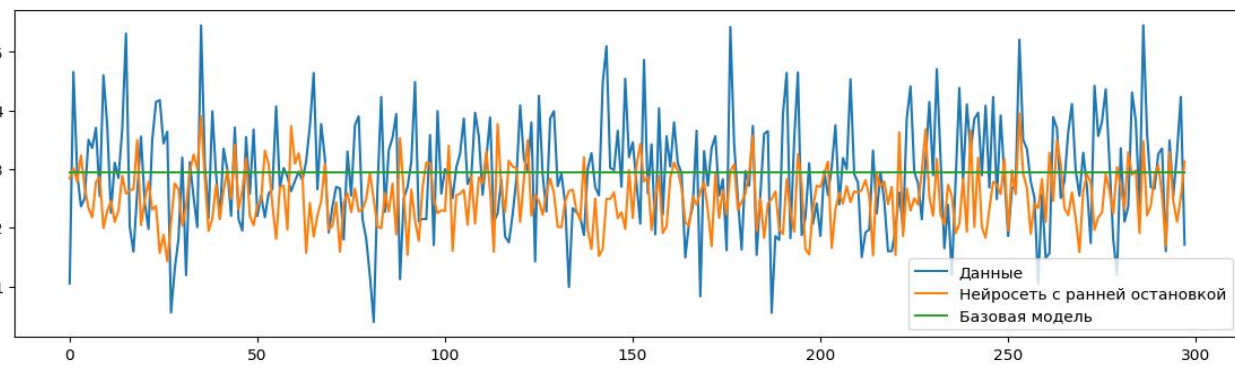


Рисунок 15 – визуализация работы модели с ранней остановкой

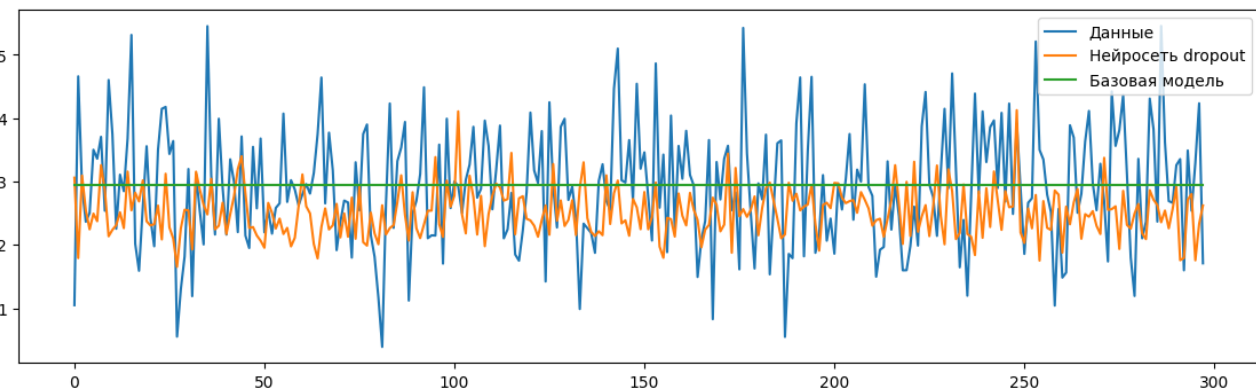


Рисунок 16 – визуализация работы нейросети Dropout



Скриншоты приложения

Прогнозирование свойств базальтопластика

Прогнозирование модуля упругости при растяжении и прочности при растяжении

Прогнозирование соотношения матрица-наполнитель

Прогнозирование модуля упругости при растяжении и прочности при растяжении

Соотношение матрица-наполнитель	2.771331
Плотность, кг/м3	2030.0
Модуль упругости, ГПа	753.0
Количество отвердителя, м. %	111.86
Содержание эпоксидных групп, %_2	22.267857
Температура вспышки, C_2	284.615385
Поверхностная плотность, г/м2	210.0
Потребление смолы, г/м2	220.0
Угол нашивки, град	0.0
Шаг нашивки	5.0
Плотность нашивки	57.0

[Предсказать](#)

Результат прогнозирования:

Модуль упругости при растяжении, ГПа 72.9901765268755	Прочность при растяжении, МПа 2437.196271717282
---	---

Прогнозирование соотношения матрица-наполнитель

Плотность, кг/м3	2030.0
Модуль упругости, ГПа	753.0
Количество отвердителя, м. %	111.86
Содержание эпоксидных групп, %_2	22.267857
Температура вспышки, C_2	284.615385
Поверхностная плотность, г/м2	210.0
Модуль упругости при растяжении, ГПа	70.0
Прочность при растяжении, МПа	3000.0
Потребление смолы, г/м2	220.0
Угол нашивки, град	0.0
Шаг нашивки	5.0
Плотность нашивки	57.0

[Предсказать](#)

Результат прогнозирования:

Соотношение матрица-наполнитель
2.8035516538318635



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru