



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

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

Прогнозирование конечных свойств новых материалов (композиционных материалов)

Докладчик: **Синцова Виктория
Викторовна**



Цель работы: создание моделей прогнозирования конечных свойств новых композиционных материалов, используя данные о начальных свойствах компонентов композиционных материалов.

Актуальность темы : созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

Предмет исследования : методы используемые в Data Science для выявления закономерностей в наборах данных.

Объект исследования : свойства композитных материалов.



Датасеты со свойствами композитов представлен в 2 файлах формата Excel.

Исходные датасеты по свойствам базальтопластика и нашивкам из углепластика были объединены по типу INNER с удалением столбцов нумерации, а количество строк приведено к наименьшему единому значению – 1023.



x_join.info()



```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 1023 entries, 0 to 1022  
Data columns (total 13 columns):
```

#	Column	Non-Null Count	Dtype
0	Соотношение матрица-наполнитель	1023 non-null	float64
1	Плотность, кг/м3	1023 non-null	float64
2	модуль упругости, ГПа	1023 non-null	float64
3	Количество отвердителя, м.%	1023 non-null	float64
4	Содержание эпоксидных групп,%_2	1023 non-null	float64
5	Температура вспышки, C_2	1023 non-null	float64
6	Поверхностная плотность, г/м2	1023 non-null	float64
7	Модуль упругости при растяжении, ГПа	1023 non-null	float64
8	Прочность при растяжении, МПа	1023 non-null	float64
9	Потребление смолы, г/м2	1023 non-null	float64
10	Угол нашивки, град	1023 non-null	int64
11	Шаг нашивки	1023 non-null	float64
12	Плотность нашивки	1023 non-null	float64

```
dtypes: float64(12), int64(1)
```

```
memory usage: 111.9 KB
```



Разведочный анализ данных

Проверка наличия пропусков

```
[ ] x_join.isnull().sum()
```

Соотношение матрица-наполнитель	0
Плотность, кг/м3	0
модуль упругости, ГПа	0
Количество отвердителя, м.%	0
Содержание эпоксидных групп,%_2	0
Температура вспышки, C_2	0
Поверхностная плотность, г/м2	0
Модуль упругости при растяжении, ГПа	0
Прочность при растяжении, МПа	0
Потребление смолы, г/м2	0
Угол нашивки, град	0
Шаг нашивки	0
Плотность нашивки	0

dtype: int64

Количество уникальных значений

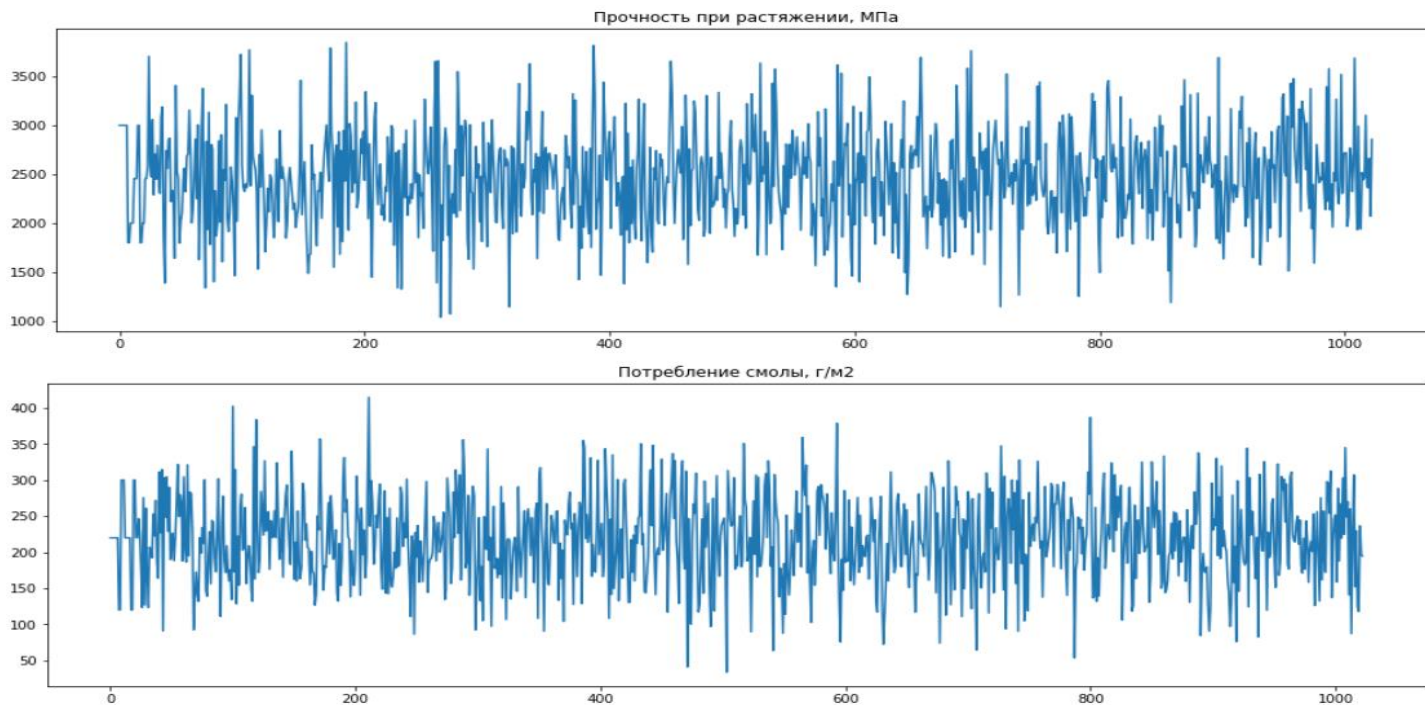
```
▶ x_join.nunique()
```

Соотношение матрица-наполнитель	1014
Плотность, кг/м3	1013
модуль упругости, ГПа	1020
Количество отвердителя, м.%	1005
Содержание эпоксидных групп,%_2	1004
Температура вспышки, C_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

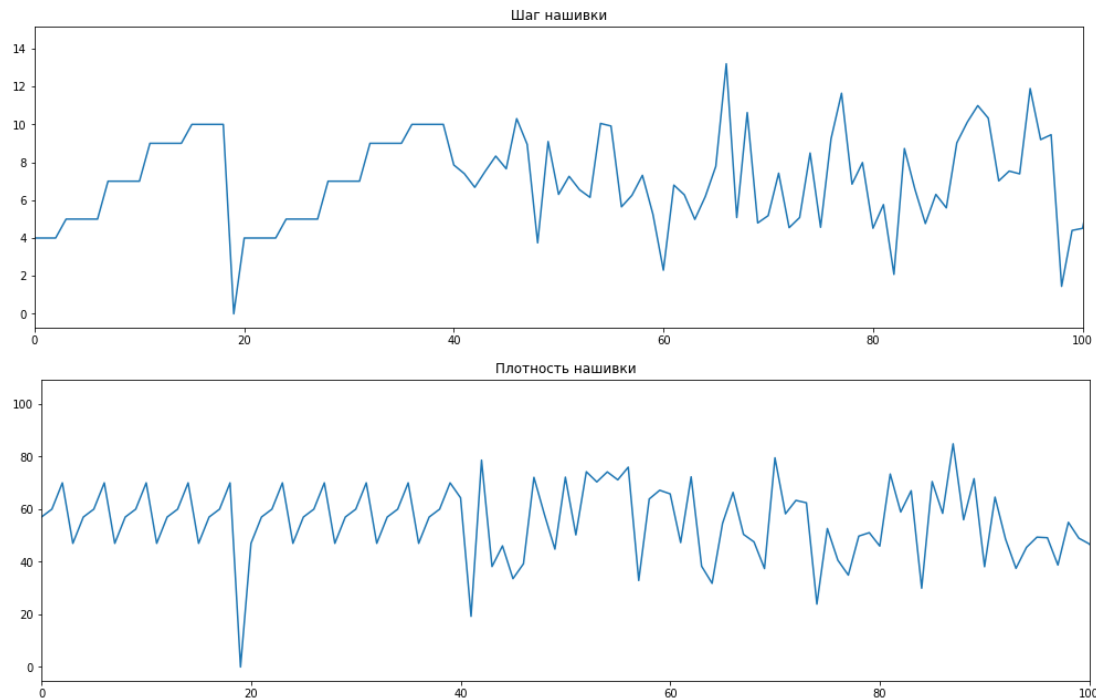


Графики данных



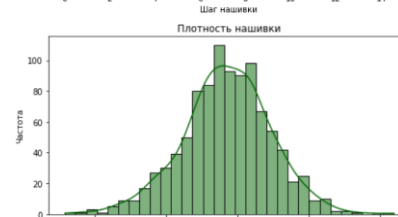
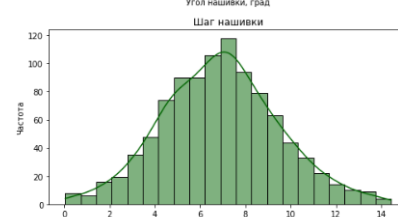
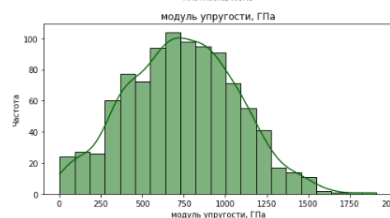
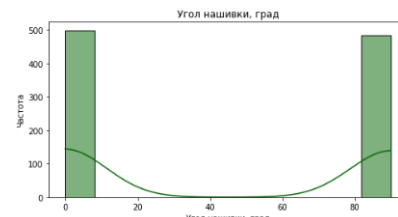
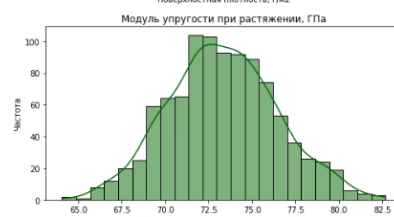
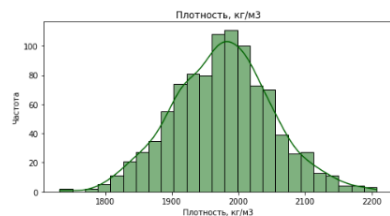
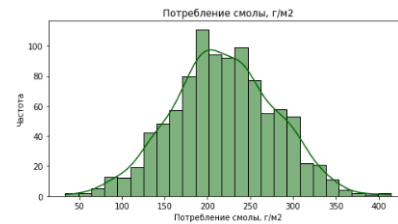
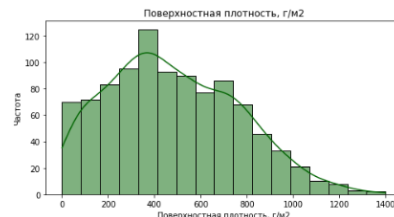


Графики данных (первые 100 значений)



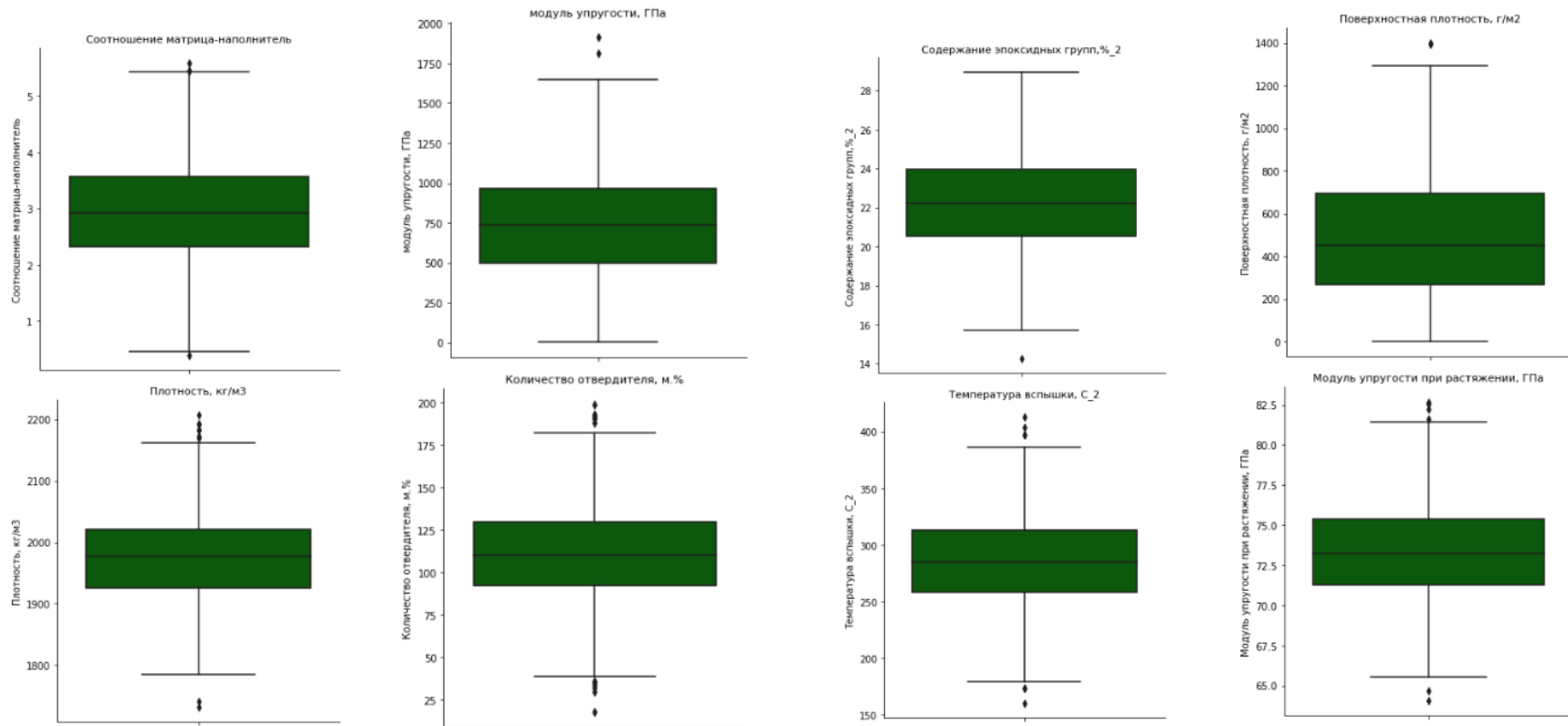


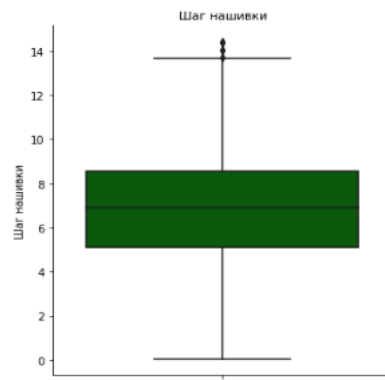
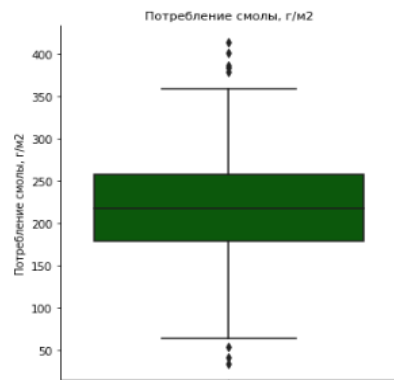
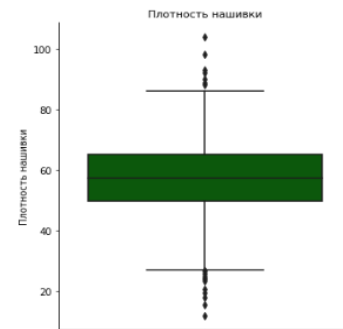
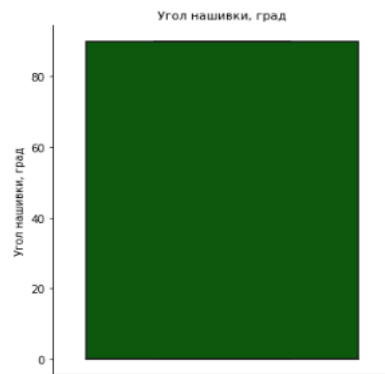
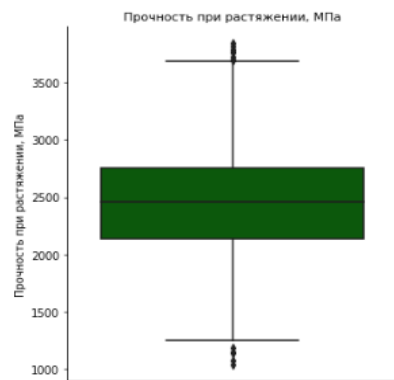
Гистограммы распределения





Диаграммы «ящик с усами»



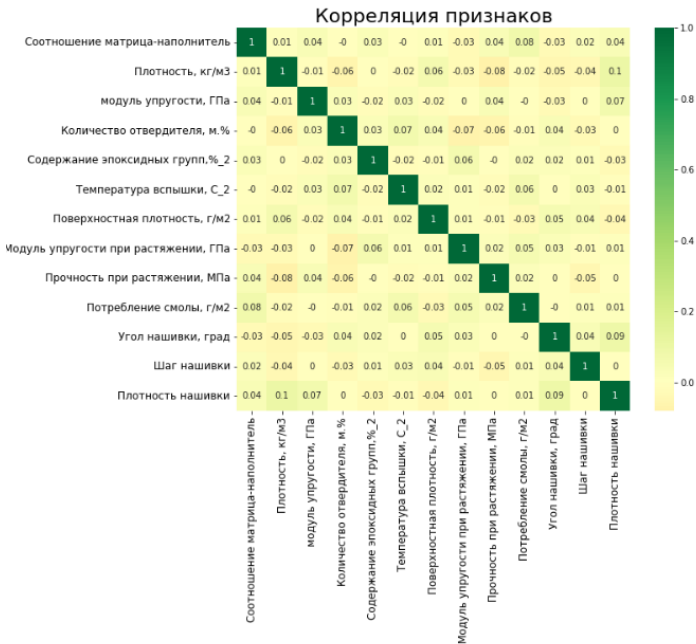




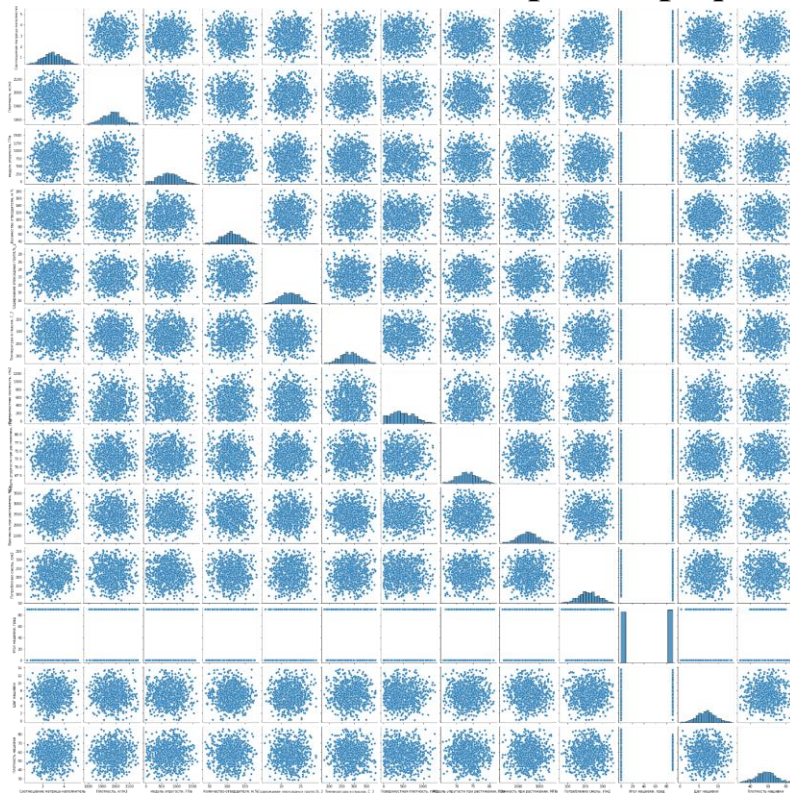
Описательная статистика

	count	mean	std	min	25%	50%	75%	max
Соотношение матрица-наполнитель	890.0	2.928925	0.896982	0.547391	2.320191	2.908811	3.551339	5.314144
Плотность, кг/м3	890.0	1974.340221	70.869622	1784.482245	1923.887189	1977.603973	2020.082671	2161.565216
модуль упругости, ГПа	890.0	735.880602	327.653047	2.436909	498.275517	733.016158	960.465724	1649.415706
Количество отвердителя, м.%	890.0	111.010058	26.973789	38.668500	92.497018	110.573604	130.404874	181.828448
Содержание эпоксидных групп,%_2	890.0	22.187011	2.418853	15.695894	20.521955	22.146953	23.966198	28.955094
Температура вспышки, C_2	890.0	285.664458	39.879163	179.374391	258.386295	285.853960	313.040444	386.067992
Поверхностная плотность, г/м2	890.0	482.105364	280.915278	0.603740	265.027350	452.891920	696.248199	1291.340115
Модуль упругости при растяжении, ГПа	890.0	73.302274	3.039914	65.793845	71.241213	73.184259	75.322715	81.417126
Прочность при растяжении, МПа	890.0	2463.507094	457.353044	1250.392802	2150.188224	2456.394188	2751.499231	3660.450210
Потребление смолы, г/м2	890.0	218.010506	57.575275	64.524180	179.719792	216.779521	256.995883	359.052220
Угол нашивки, град	890.0	45.910112	45.016093	0.000000	0.000000	90.000000	90.000000	90.000000
Шаг нашивки	890.0	6.909465	2.506636	0.145034	5.167202	6.922196	8.568608	13.653987
Плотность нашивки	890.0	57.428007	11.288299	28.237746	50.211993	57.546947	64.798593	86.012427

Тепловая карта коэффициентов корреляции



Попарные графики рассеяния точек



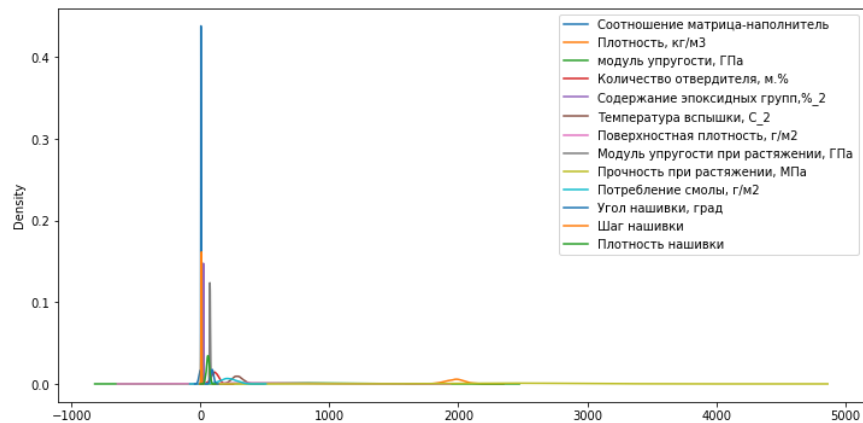
Разведочный анализ данных показал, что линейной связи между любыми переменными нет, корреляция равна 0.

При этом все параметры (за исключением параметра «Угол нашивки») имеют нормальное распределение, что может свидетельствовать о случайной генерации чисел.

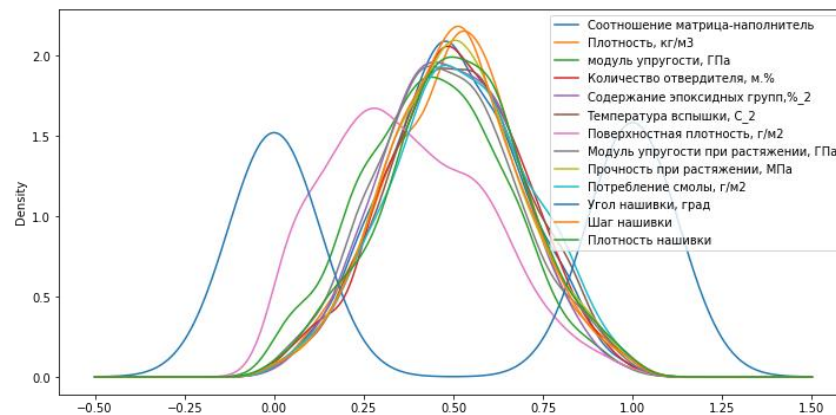


Предобработка данных

Разброс данных до нормализации



Разброс данных после нормализации





Разработка и обучение модели

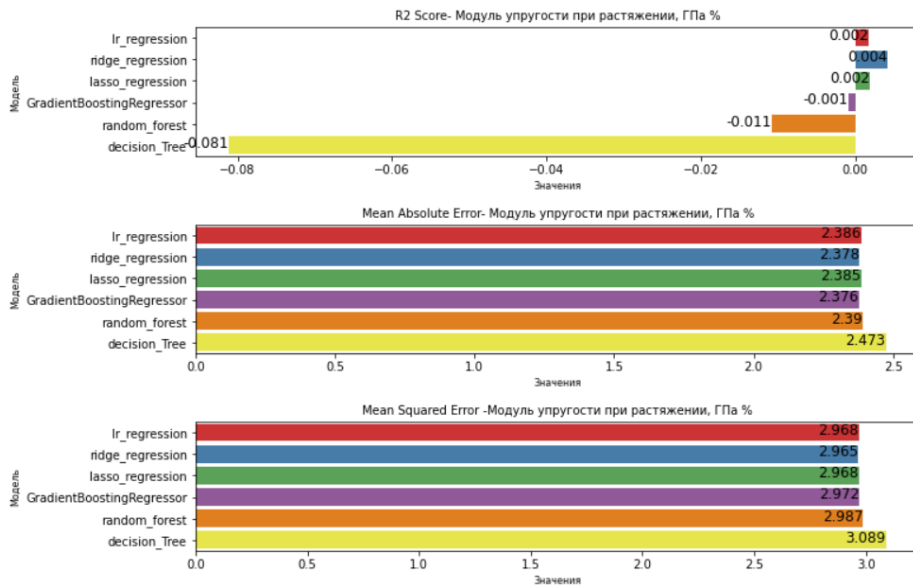
Для решения задачи предсказания модуля упругости при растяжении и прочности при растяжении были использованы следующие методы:

- линейная регрессия (Linear Regression);
- гребневая регрессия (Ridge Regression);
- регрессия по методу наименьших квадратов (Lasso Regression);
- градиентный бустинг (Gradient Boosting);
- случайные лес (Random Forest Regression);
- регрессия дерева решений (Decision Tree Regression).

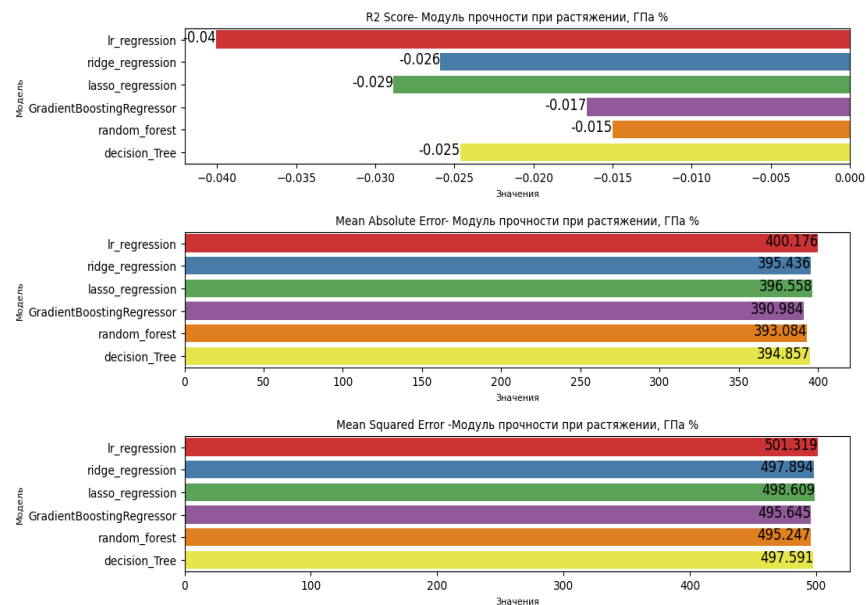
В качестве оценки работы моделей были использованы следующие метрики: MSE (среднеквадратичная ошибка), R^2 (Коэффициент детерминации), MAE (средняя абсолютная ошибка).

Результаты оценок каждой модели

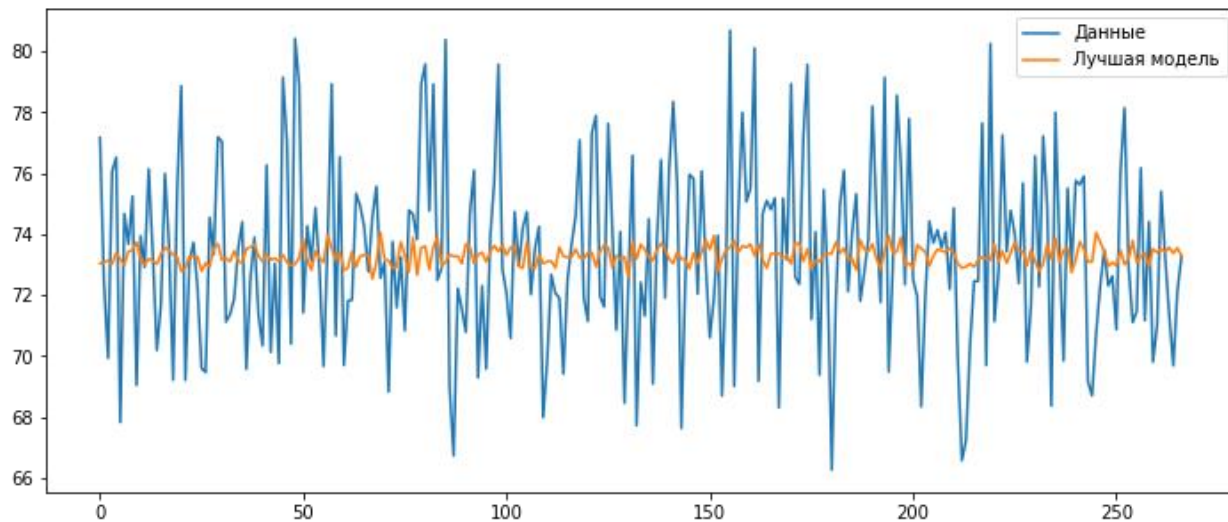
Ошибки модели предсказания модуля упругости



Ошибки модели предсказания прочности

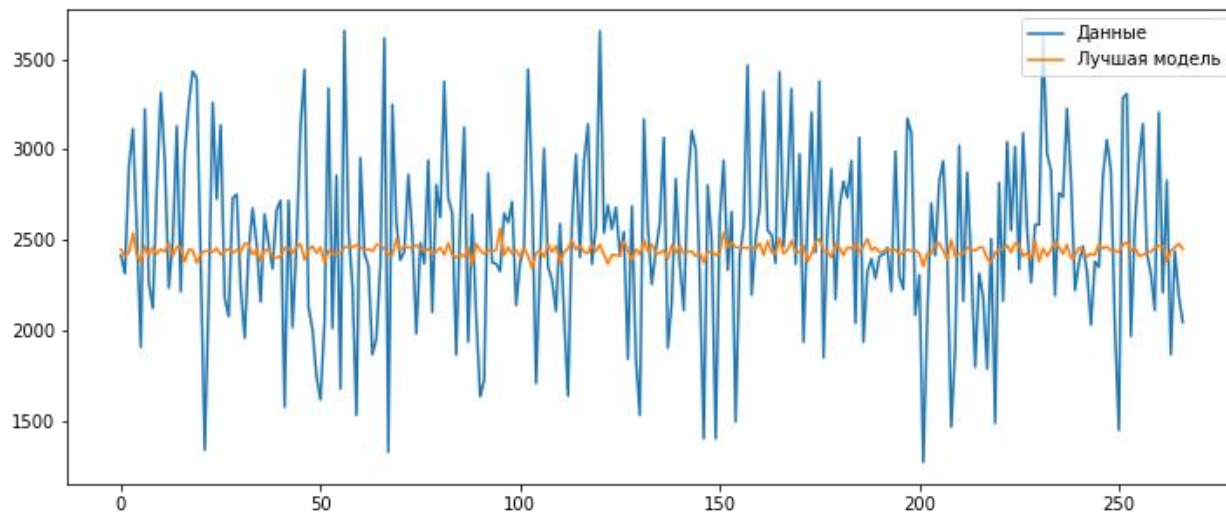


Визуализация работы лучшей модели



Модуль упругости при растяжении

Визуализация работы лучшей модели



Прочность при растяжении



Разработка нейронной сети для прогнозирования

Model: "sequential_5"

Layer (type)	Output Shape	Param #
dense_24 (Dense)	(None, 50)	550
dropout_19 (Dropout)	(None, 50)	0
dense_25 (Dense)	(None, 128)	6528
dropout_20 (Dropout)	(None, 128)	0
dense_26 (Dense)	(None, 64)	8256
dropout_21 (Dropout)	(None, 64)	0
dense_27 (Dense)	(None, 1)	65

=====
Total params: 15,399
Trainable params: 15,399
Non-trainable params: 0

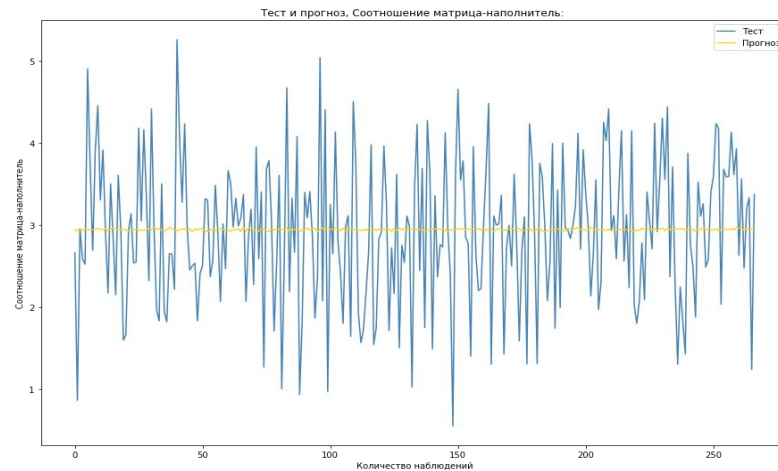
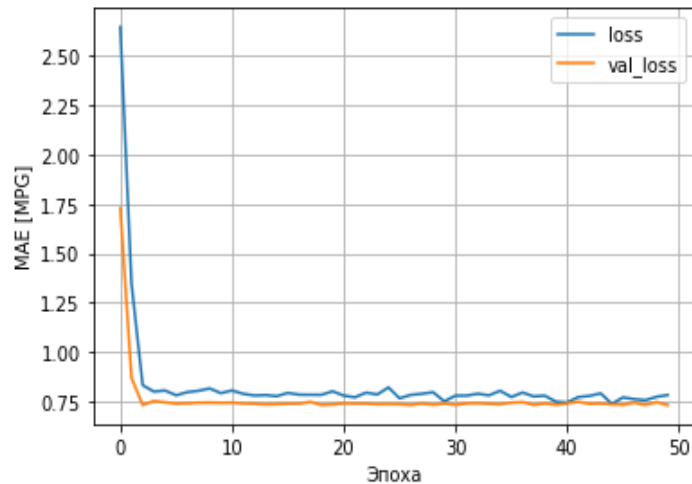
Архитектура нейронной сети:

- 2 скрытых уровня (первый содержит 128 нейрона, второй – 64 нейрона);
- во входном слое 10 признаков;
- выходной слой с 1 нейроном;
- активационная функция скрытых слоев – softplus.
- для оптимизации был применен метод Adam.
- Для борьбы с переобучением добавлены Dropout-слои с параметром 0.12.

Обучение нейросети происходило со следующими параметрами:

- пропорция разбиения данных на тестовые и валидационные: 20%;
- количество эпох - 100.

Визуализация результатов работы нейросети



Ошибки модели:

- $MSE = 0,783$;
- $R^2 = -0,0007$



Разработка приложения



Предсказание значений модуля упругости при растяжении и прочности при растяжении

Введите значения известных параметров:

Плотность, кг/м^3 =

Модуль упругости, ГПа =

Количество отвердителя, м.%% =

Содержание эпоксидных групп, %_2 =

Температура вспышки, C_{-2} =

Поверхностная плотность, г/м^2 =

Прочность при растяжении, МПа =

Потребление смолы, г/м^2 =

Угол нашивки, град =

Шаг нашивки =

Результат модели:

Модуль упругости при растяжении, ГПа =

Прочность при растяжении, МПа =



Заключение

В ходе решения задачи прогнозирования конечных свойств новых материалов были изучены основные теоретические и практические методы машинного обучения. Проведен предварительный анализ данных и их предобработка. Изучены основные алгоритмы машинного обучения и проведен сравнительный анализ полученных результатов. В моделях были настроены гиперпараметры.

После выполнения исследования разработано веб-приложение, данные загружены в репозиторий.

В ходе выполнения ВКР не удалось разработать модель, которая предсказывала бы значения с приемлемой точностью. Модель нейронной сети так же не показала приемлемого результата.



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

Спасибо за внимание!