

Образовательный центр МГТУ им. Н.Э.
Баумана
Выпускная квалификационная работа по
курсу "Data Science»

Тема: Прогнозирование конечных свойств новых материалов
(композиционных материалов)

Слушатель: Урушадзе Тамара

Постановка задачи

- изучить предметную область
- провести разведочный анализ данных
- разделить данные на тренировочную и тестовую выборки
- выполнить препроцессинг (предобработку)
- выбрать базовую модель и модели для подбора
- подобрать гиперпараметры с помощью поиска по сетке с перекрестной проверкой
- сравнить модели после подбора гиперпараметров и выбрать лучшую
- сравнить качество лучшей и базовой моделей на тестовой выборке
- сравнить качество лучшей модели на тренировочной и тестовой выборке
- разработать приложение

Разведочный анализ данных

- X_bp (матрица из базальтопластика):
признаков: 10 и индекс строк: 1023
- X_nip (наполнитель из углепластика):
признаков: 3 и индекс строк: 1040

Объединение с типом INNER по индексу, получилось:

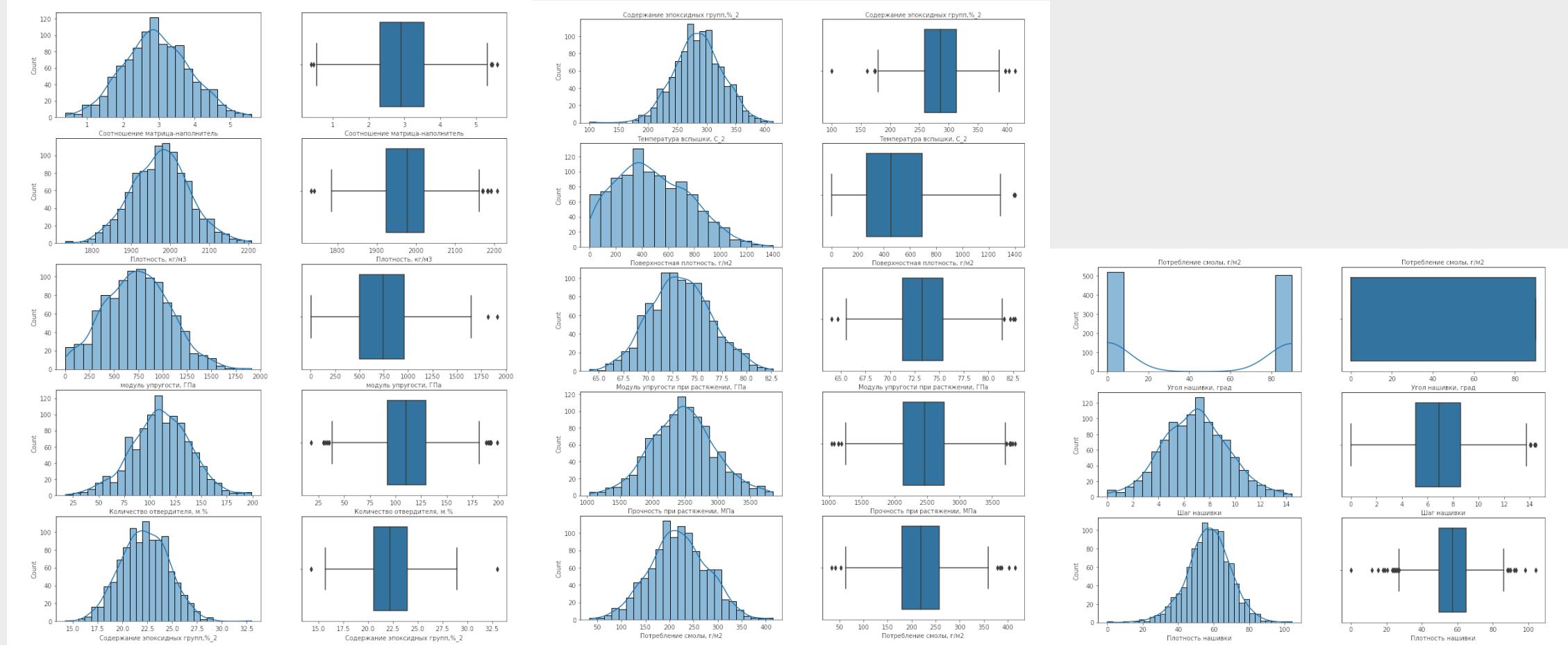
признаков: 13 строк: 1023

Разведочный анализ данных

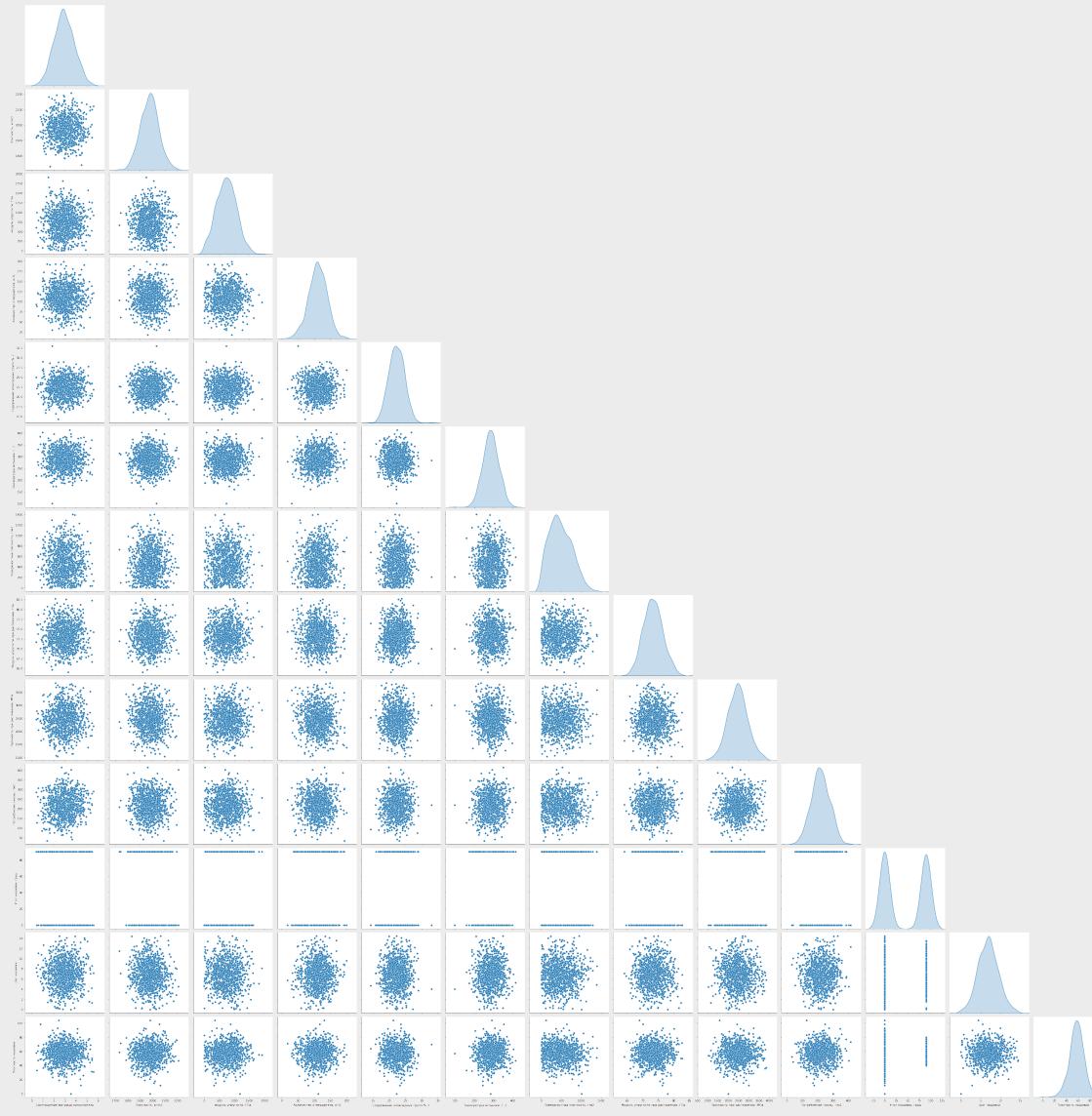
| Название | Файл | Тип данных | Непустых значений | Уникальных значений |
|--------------------------------------|-------|------------|-------------------|---------------------|
| Соотношение матрица-наполнитель | X_bp | float64 | 1023 | 1014 |
| Плотность, кг/м3 | X_bp | float64 | 1023 | 1013 |
| модуль упругости, ГПа | X_bp | float64 | 1023 | 1020 |
| Количество отвердителя, м.% | X_bp | float64 | 1023 | 1005 |
| Содержание эпоксидных групп,%_2 | X_bp | float64 | 1023 | 1004 |
| Температура вспышки, С_2 | X_bp | float64 | 1023 | 1003 |
| Поверхностная плотность, г/м2 | X_bp | float64 | 1023 | 1004 |
| Модуль упругости при растяжении, ГПа | X_bp | float64 | 1023 | 1004 |
| Прочность при растяжении, МПа | X_bp | float64 | 1023 | 1004 |
| Потребление смолы, г/м2 | X_bp | float64 | 1023 | 1003 |
| Угол нашивки, град | X_nip | float64 | 1023 | 2 |
| Шаг нашивки | X_nip | float64 | 1023 | 989 |
| Плотность нашивки | X_nip | float64 | 1023 | 988 |

| | Среднее | Стандартное отклонение | Минимум | Максимум | Медиана |
|--------------------------------------|-----------|------------------------|-----------|-----------|-----------|
| Соотношение матрица-наполнитель | 2.9304 | 0.9132 | 0.3894 | 5.5917 | 2.9069 |
| Плотность, кг/м3 | 1975.7349 | 73.7292 | 1731.7646 | 2207.7735 | 1977.6217 |
| модуль упругости, ГПа | 739.9232 | 330.2316 | 2.4369 | 1911.5365 | 739.6643 |
| Количество отвердителя, м.% | 110.5708 | 28.2959 | 17.7403 | 198.9532 | 110.5648 |
| Содержание эпоксидных групп, %_2 | 22.2444 | 2.4063 | 14.2550 | 33.0000 | 22.2307 |
| Температура вспышки, С_2 | 285.8822 | 40.9433 | 100.0000 | 413.2734 | 285.8968 |
| Поверхностная плотность, г/м2 | 482.7318 | 281.3147 | 0.6037 | 1399.5424 | 451.8644 |
| Модуль упругости при растяжении, ГПа | 73.3286 | 3.1190 | 64.0541 | 82.6821 | 73.2688 |
| Прочность при растяжении, МПа | 2466.9228 | 485.6280 | 1036.8566 | 3848.4367 | 2459.5245 |
| Потребление смолы, г/м2 | 218.4231 | 59.7359 | 33.8030 | 414.5906 | 219.1989 |
| Угол нашивки, град | 44.2522 | 45.0158 | 0.0000 | 90.0000 | 0.0000 |
| Шаг нашивки | 6.8992 | 2.5635 | 0.0000 | 14.4405 | 6.9161 |
| Плотность нашивки | 57.1539 | 12.3510 | 0.0000 | 103.9889 | 57.3419 |

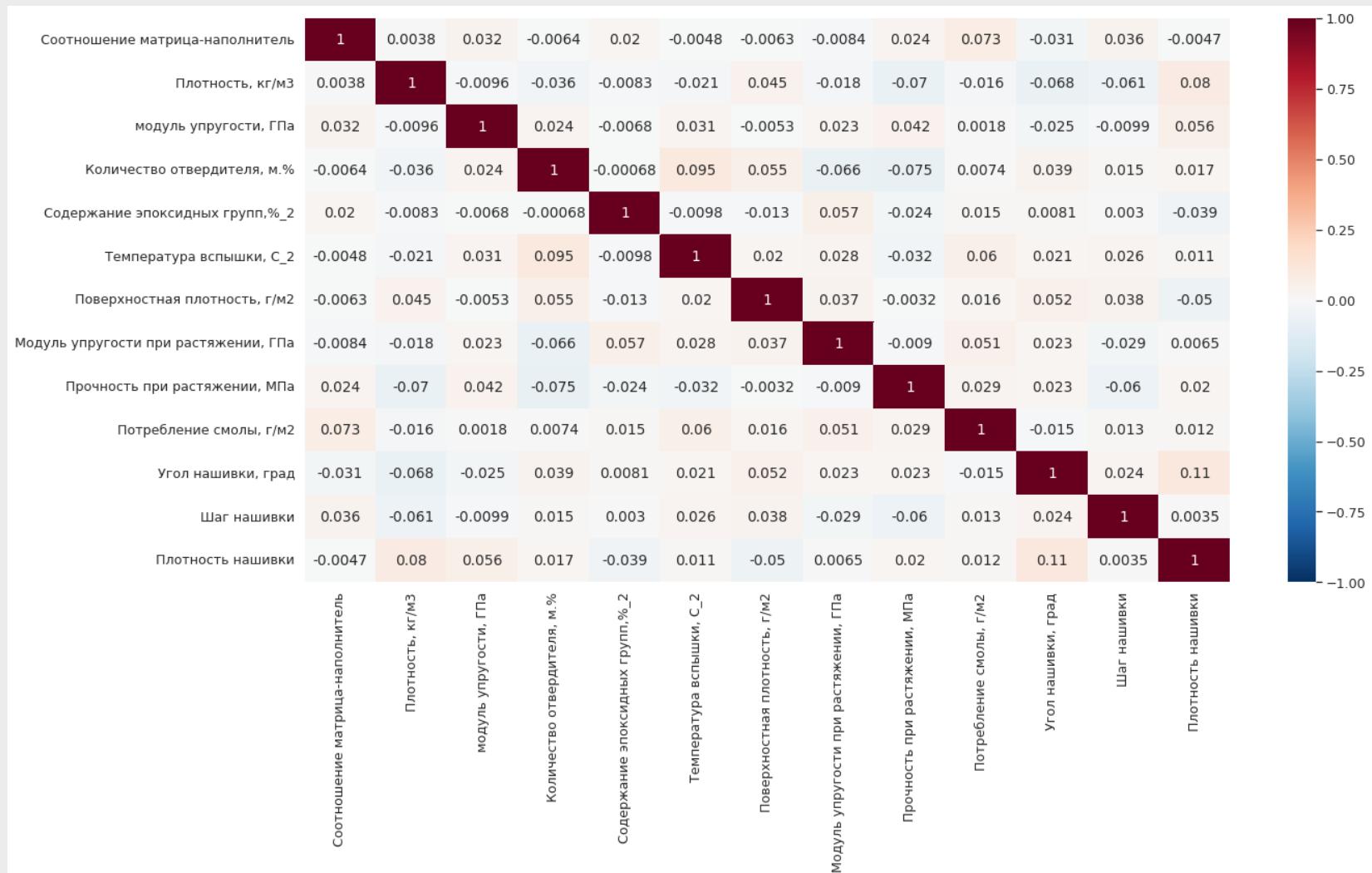
Гистограммы распределения и диаграммы “ящик с усами”



Попарные графики рассеяния точек



Матрица корреляции



Метрики качества

- R² или коэффициент детерминации
- RMSE (Root Mean Squared Error) или корень из средней квадратичной ошибки
- MAE (Mean Absolute Error) или средняя абсолютная ошибка
- MAPE (Mean Absolute Percentage Error) или средняя абсолютная процентная ошибка
- max error или максимальная ошибка данной модели

Модели

- Линейная регрессия
- Лассо (LASSO) и гребневая (Ridge) регрессия
- Метод опорных векторов для регрессии
- Метод k-ближайших соседей
- Деревья решений
- Случайный лес
- Градиентный бустинг (GradientBoostingRegressor, LGBMRegressor, CatBoostRegressor)
- Нейронная сеть

Модель для модуля упругости при растяжении

| | R2 | RMSE | MAE | MAPE | max_error |
|---------------------------|-----------|----------|----------|----------|-----------|
| DummyRegressor | -0.000637 | 3.037614 | 2.451541 | 0.033474 | 9.184598 |
| LinearRegression | -0.027256 | 3.077751 | 2.490270 | 0.034001 | 8.696302 |
| Ridge | -0.008641 | 3.049738 | 2.462914 | 0.033627 | 8.950818 |
| Lasso | -0.013971 | 3.057785 | 2.470518 | 0.033731 | 8.845795 |
| SVR | -0.003542 | 3.042020 | 2.453717 | 0.033486 | 9.134810 |
| KNeighborsRegressor | -0.043995 | 3.102726 | 2.514148 | 0.034338 | 8.648469 |
| DecisionTreeRegressor | -0.007256 | 3.047643 | 2.468863 | 0.033730 | 9.141288 |
| RandomForestRegressor | -0.002240 | 3.040045 | 2.450632 | 0.033464 | 9.021327 |
| CatBoostRegressor | -0.153545 | 3.261455 | 2.575692 | 0.035115 | 10.221554 |
| LGBMRegressor | -0.131812 | 3.230586 | 2.588977 | 0.035375 | 9.272518 |
| GradientBoostingRegressor | -0.010664 | 3.052795 | 2.471146 | 0.033743 | 9.001461 |

Модель для модуля упругости при растяжении



```
{'rfr__max_depth': 3, 'rfr__max_features': 1, 'rfr__n_estimators': 150}  
0.0007388995487566441
```

Модель для прочности при растяжении

| | R2 | RMSE | MAE | MAPE | max_error |
|---------------------------|-----------|------------|------------|----------|-------------|
| DummyRegressor | -0.000126 | 486.661114 | 377.888006 | 0.169870 | 1429.643384 |
| LinearRegression | -0.034072 | 494.851306 | 390.849482 | 0.173599 | 1400.304664 |
| Ridge | -0.013243 | 489.842094 | 384.184501 | 0.171572 | 1400.675910 |
| Lasso | -0.025234 | 492.732135 | 388.599744 | 0.172956 | 1386.105483 |
| SVR | -0.000684 | 486.796951 | 378.563890 | 0.169446 | 1417.964939 |
| KNeighborsRegressor | -0.023321 | 492.272110 | 386.349169 | 0.172472 | 1462.048774 |
| DecisionTreeRegressor | -0.022080 | 491.973684 | 381.454292 | 0.171453 | 1410.035042 |
| RandomForestRegressor | -0.006850 | 488.294456 | 381.044142 | 0.170590 | 1446.797333 |
| CatBoostRegressor | -0.211601 | 535.647683 | 422.471777 | 0.187321 | 1626.489884 |
| LGBMRegressor | -0.201972 | 533.514941 | 424.111028 | 0.188525 | 1638.869758 |
| GradientBoostingRegressor | -0.016042 | 490.518418 | 384.057211 | 0.171454 | 1410.593198 |

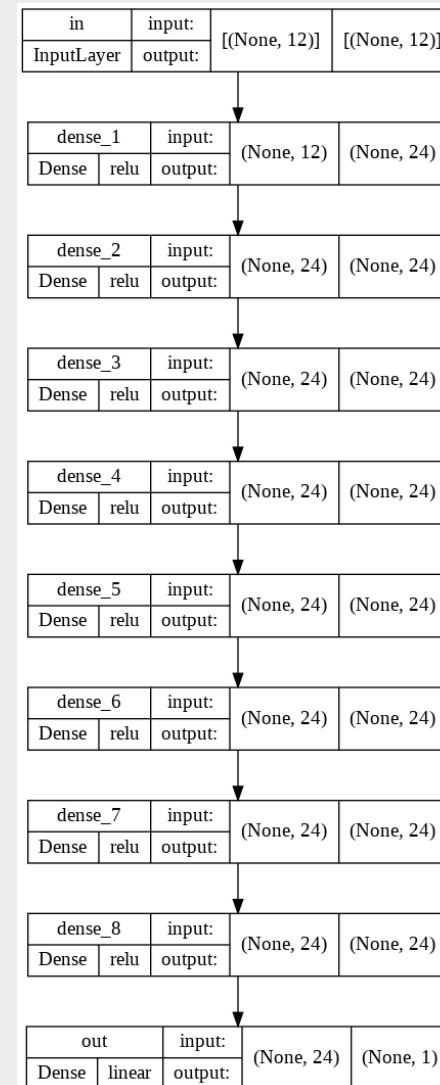
Модель для прочности при растяжении



```
{'svr_C': 0.02, 'svr_kernel': 'poly'}  
-0.024098200507591927
```

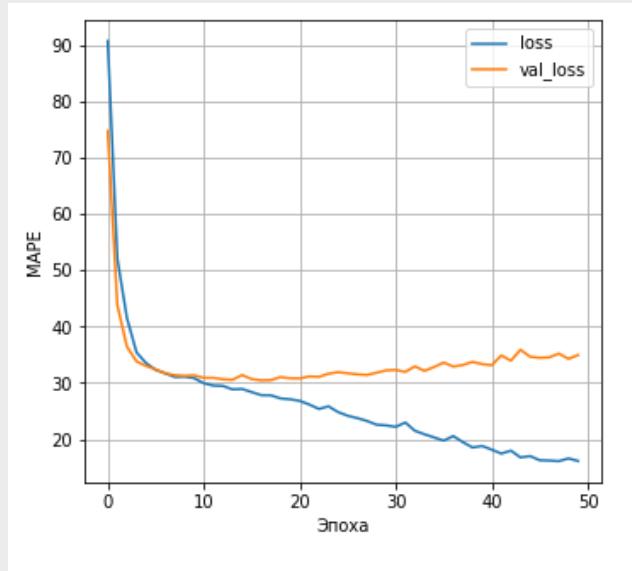
Модель для соотношения матрица-наполнитель

| Layer (type) | Output Shape | Param # |
|-------------------------|--------------|---------|
| <hr/> | | |
| dense_1 (Dense) | (None, 24) | 312 |
| dense_2 (Dense) | (None, 24) | 600 |
| dense_3 (Dense) | (None, 24) | 600 |
| dense_4 (Dense) | (None, 24) | 600 |
| dense_5 (Dense) | (None, 24) | 600 |
| dense_6 (Dense) | (None, 24) | 600 |
| dense_7 (Dense) | (None, 24) | 600 |
| dense_8 (Dense) | (None, 24) | 600 |
| out (Dense) | (None, 1) | 25 |
| <hr/> | | |
| Total params: 4,537 | | |
| Trainable params: 4,537 | | |
| Non-trainable params: 0 | | |

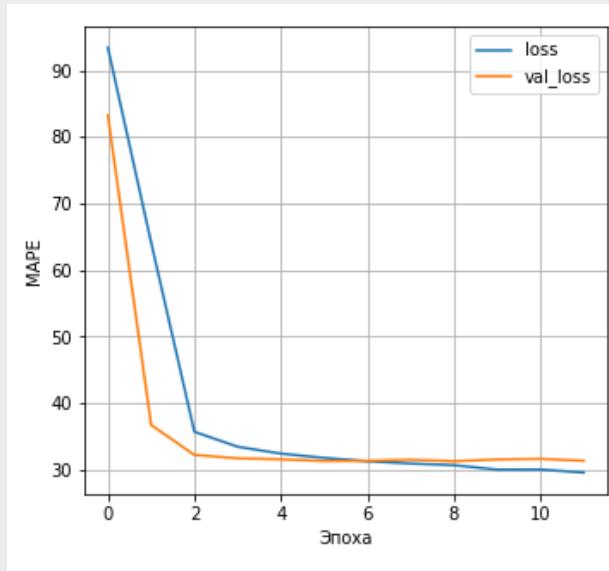


Графики обучения нейросети

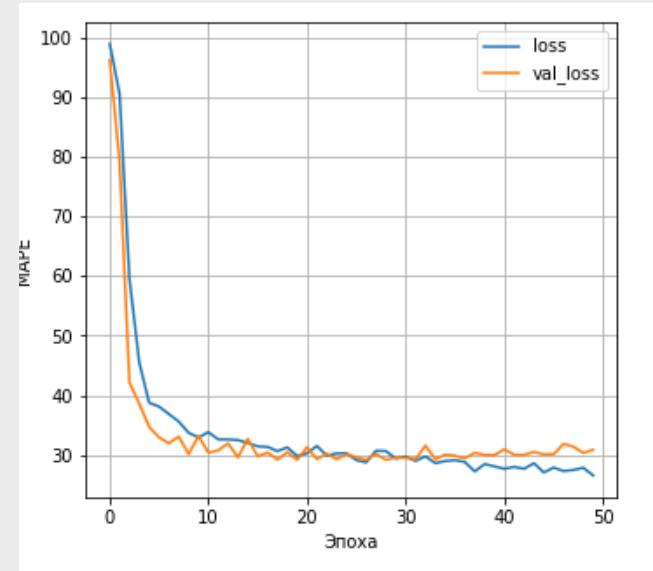
Обучение
базовой
нейросети



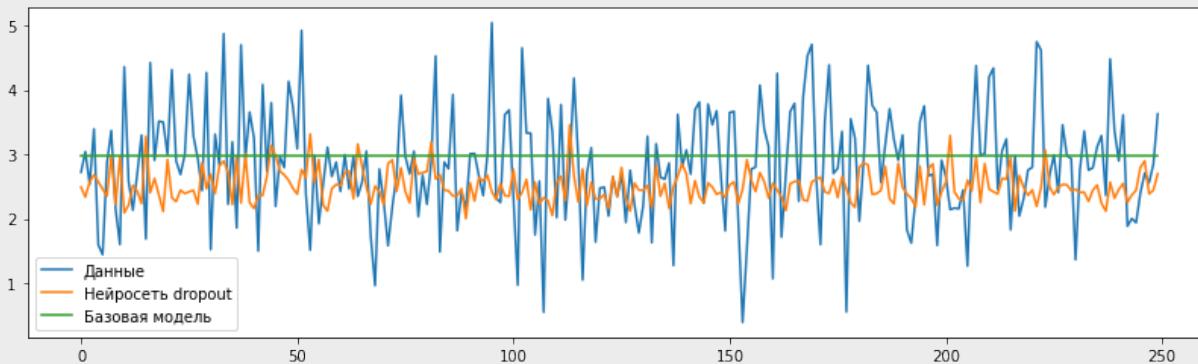
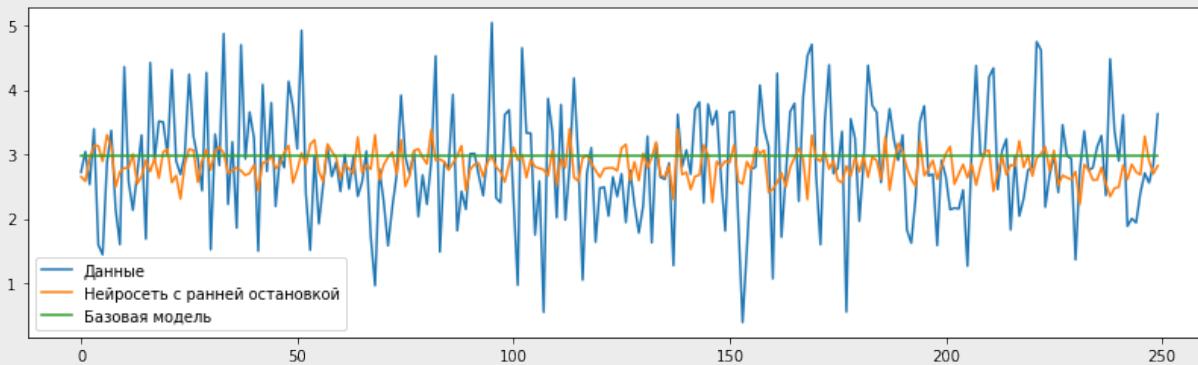
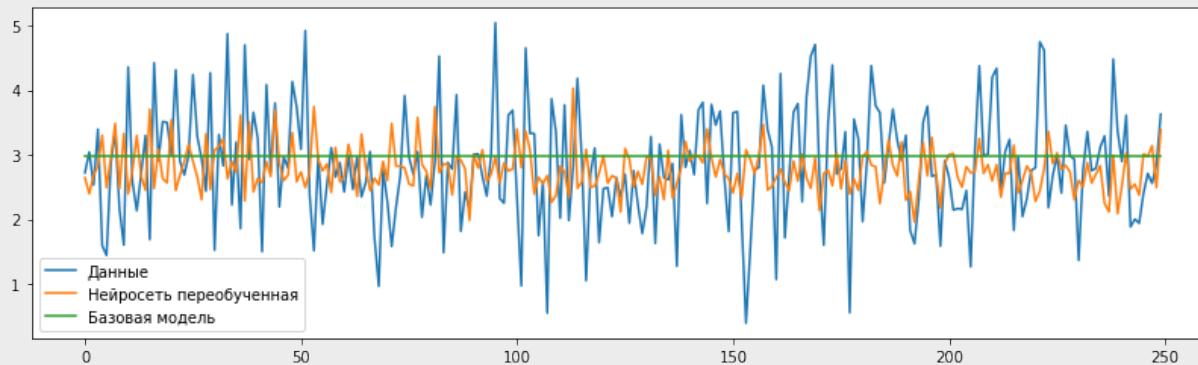
Борьба с
переобучением:
ранняя остановка



Борьба с
переобучением:
Dropout-слои



Визуализация результатов работы нейросетей



Метрики работы нейросетей на тестовом множестве

| | R2 | RMSE | MAE | MAPE | max_error |
|-------------------------------|-----------|----------|----------|----------|-----------|
| DummyRegressor | -0.020151 | 0.882319 | 0.686249 | 0.341972 | 2.577898 |
| Нейросеть переобученная | -0.257839 | 0.979728 | 0.775280 | 0.318282 | 2.679286 |
| Нейросеть с ранней остановкой | -0.030224 | 0.886664 | 0.688314 | 0.323383 | 2.342318 |
| Нейросеть dropout | -0.274153 | 0.986061 | 0.779692 | 0.324875 | 2.624417 |

Разработка веб-приложения

Выпускная квалификационная работа по курсу «Data Science»
Прогнозирование конечных свойств новых материалов (композиционных материалов)
Студентка Тамара Урушадзе

[Прогнозирование модуля упругости при растяжении и прочности при растяжении](#)

Прогнозирование модуля упругости при растяжении и прочности при растяжении

| | |
|--|---|
| <input type="text"/> | Соотношение матрица-наполнитель |
| <input type="text"/> | Плотность, кг/м ³ |
| <input type="text"/> | Модуль упругости, ГПа |
| <input type="text"/> | Количество отвердителя, м.% |
| <input type="text"/> | Содержание эпоксидных групп, %_2 |
| <input type="text"/> | Температура вспышки, С_2 |
| <input type="text"/> | Поверхностная плотность, г/м ² |
| <input type="text"/> | Потребление смолы, г/м ² |
| <input type="text"/> | Угол нашивки, град |
| <input type="text"/> | Шаг нашивки |
| <input type="text"/> | Плотность нашивки |
| <input type="button" value="Отправить"/> | |

Стартовая страница

Ввод входных параметров
для прогнозирования

Разработка веб-приложения

Прогнозирование модуля упругости при растяжении и прочности при растяжении

Содержание эпоксидных групп,%_2 – не должно быть отрицательным числом
Потребление смолы, г/м2 – could not convert string to float: 'test'
Угол нашивки, град – должно быть 0 или 90

Соотношение матрица-наполнитель
 Плотность, кг/м3
 Модуль упругости, ГПа
 Количество отвердителя, м.%
 Содержание эпоксидных групп,%_2
 Температура вспышки, С_2
 Поверхностная плотность, г/м2
 Потребление смолы, г/м2
 Угол нашивки, град
 Шаг нашивки
 Плотность нашивки

Прогнозирование модуля упругости при растяжении и прочности при растяжении

Соотношение матрица-наполнитель
 Плотность, кг/м3
 Модуль упругости, ГПа
 Количество отвердителя, м.%
 Содержание эпоксидных групп,%_2
 Температура вспышки, С_2
 Поверхностная плотность, г/м2
 Потребление смолы, г/м2
 Угол нашивки, град
 Шаг нашивки
 Плотность нашивки

Входные данные:

| Соотношение матрица-наполнитель | Плотность, кг/м3 | модуль упругости, ГПа | Количество отвердителя, м.% | Содержание эпоксидных групп,%_2 | Температура вспышки, С_2 | Поверхностная плотность, г/м2 | Потребление смолы, г/м2 | Угол нашивки, град | Шаг нашивки | Плотность нашивки |
|---------------------------------|------------------|-----------------------|-----------------------------|---------------------------------|--------------------------|-------------------------------|-------------------------|--------------------|-------------|-------------------|
| 0 1.857143 | 2030.0 | 738.736842 | 50.0 | 23.75 | 284.615385 | 210.0 | 220.0 | 0.0 | 4.0 | 60.0 |

Результат модели:

Модуль упругости при растяжении Прочность при растяжении
73.36696021503352 2510.8182638288736

Проверка входных параметров

Результат работы модели

Спасибо за внимание!



edu.bmstu.ru

+7 495 182-83-85

edu@bmstu.ru

Москва, Госпитальный переулок , д.
4-6, с.3