



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

по курсу  
“Data Science 2022 4.0”.

Способы прогнозирования конечных свойств новых композиционных материалов и разработка моделей для выполнения прогнозов.

Куриной Александр  
Васильевич.



# [ Постановка задачи. ]

**Цель** выпускной квалификационной работы:

изучение способов прогнозирования конечных свойств новых композиционных материалов и разработка моделей для выполнения прогнозов.

Для достижения данной цели необходимо **решение следующих Задач:**

- **разработка алгоритма машинного обучения** для прогноза значений модуля упругости при растяжении и прочности при растяжении;
- **разработка нейронной сети** для рекомендации соотношения «матрица-наполнитель»;

1

- Для задачи прогнозирования модуля упругости при растяжении и прочности при растяжении планируется рассмотреть следующие модели:
  - Метод K-ближайших соседей;
  - Линейная регрессия;
  - AdaBoost;
  - XGBoost;
  - Стохастический градиентный спуск;
  - Метод опорных векторов;
  - Дерево принятия решений;
  - Градиентный бустинг;
  - Случайный лес;
  - Метод регрессии «Lasso»;
- Для построения рекомендательной нейронной сети планируется воспользоваться поиском по сетке с перекрестной проверкой;



# [ Прделанные работы. ]

## Проведены следующие работы:

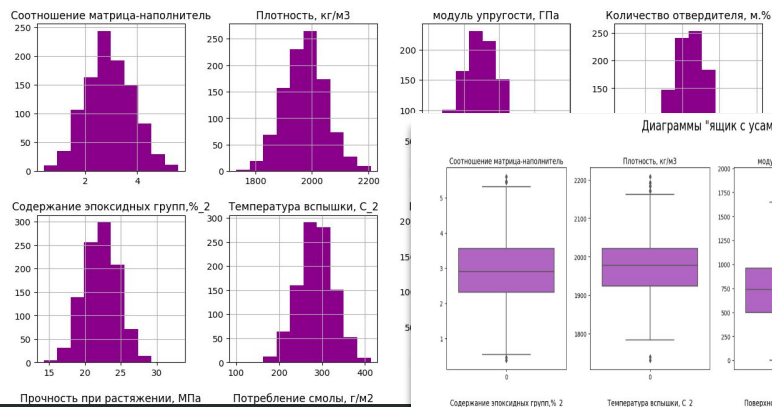
- Первичный анализ исходного датасета, объединение файлов (inner).  
/ `df.shape(); df.info(); df.isna().sum(); df.nunique(); df.describe(); pd.merge()` тип 'inner';
- Разведочный анализ предложенных данных:
  - Анализ неподготовленных данных (профайлинг pandas/ipython, три метода оценки корреляции, анализ выбросов) и принятие решения о дальнейшей предобработке данных;  
/ `ydata-profiling(); ProfileReport(); df.corr() method='pearson/spearman/kendall'; df.mean(); df.median();`
  - Визуальная оценка (гистограммы распределения каждой из переменной, диаграммы “ящик с усами”, попарные графики рассеяния точек, тепловая карта);
- Предобработка данных и их повторный анализ (работа с шумами и выбросами, нормализация и стандартизация данных).  
/ `z-оценка ( $z = (X - \mu) / \sigma$ ); IQR; sns.kdeplot(); scipy.stats.shapiro(); df.hist(); QQ-графики; MinMaxScaler(); RobustScaler(); StandardScaler(); Normalizer();`
- Обучение и оценка точности ряда моделей для прогноза модуля упругости при растяжении и прочности при растяжении.  
/ 1. Метод К-ближайших соседей; 2. Метод опорных векторов; 3. Линейная регрессия; 4. Дерево решений; 5. AdaBoost; 6. Градиентный бустинг; 7. XGBoost; 8. Случайный лес; 9. Стохастический градиентный спуск; 10. Метод регрессии «Lasso»;
- Создание и оценка точности рекомендательной нейронной сети, по параметрам соотношения «матрица-наполнитель».
- Разработка приложения (flask/console), прогнозирующего соотношение «матрица-наполнитель».



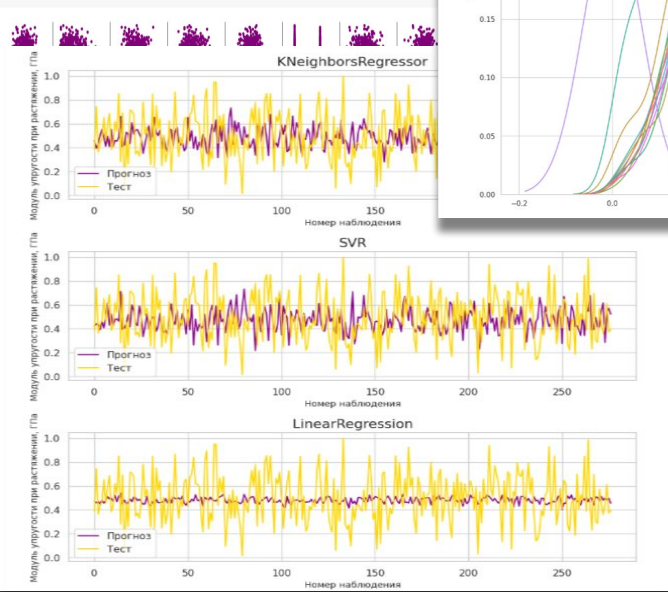
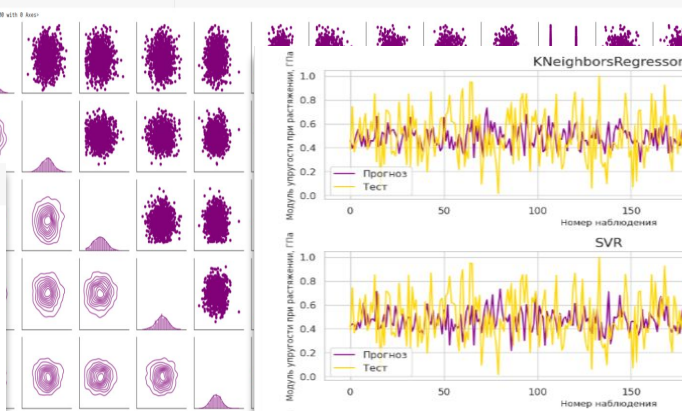
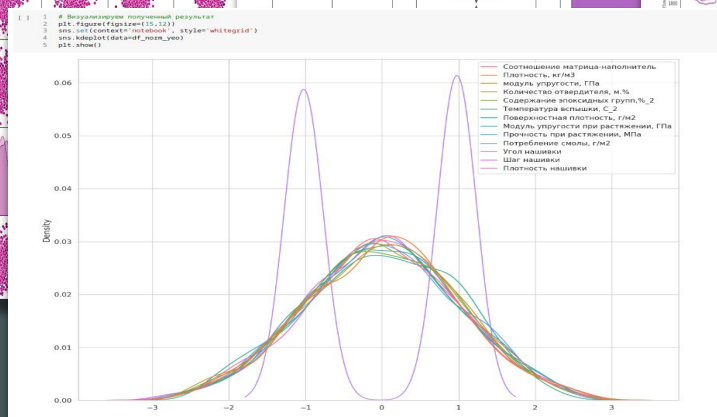
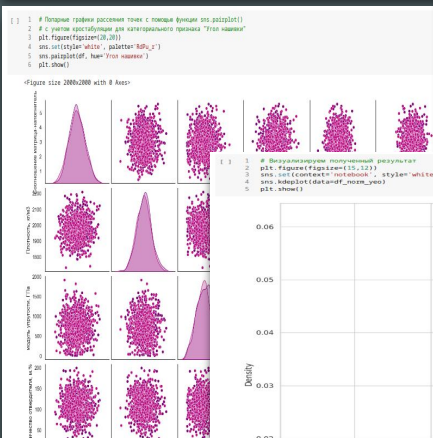
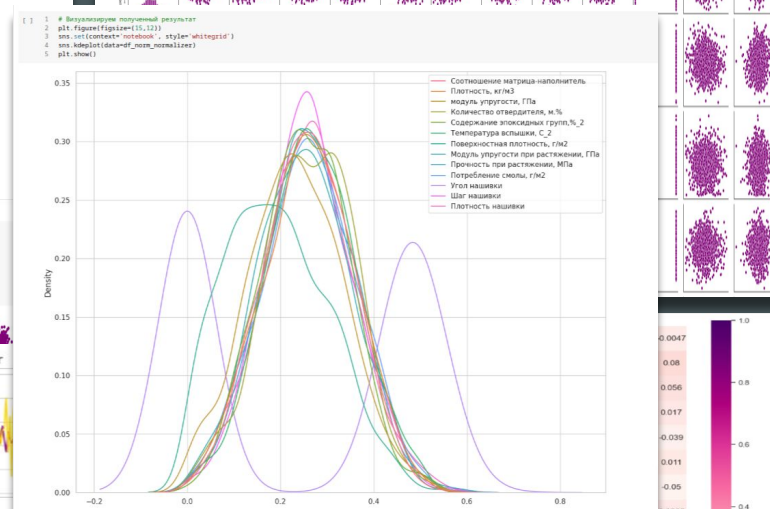
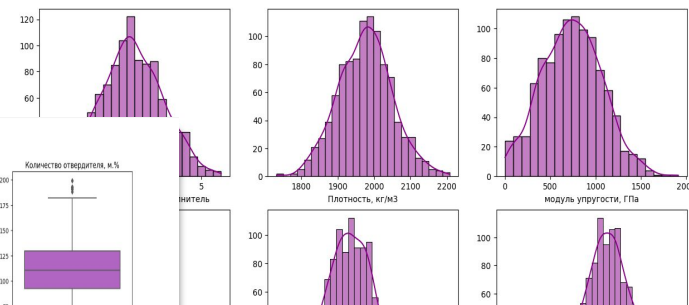


# Ход работ.

Гистограммы распределения



Гистограммы и графики плотности



факторы, МПа	0.024	-0.07	0.042	-0.075	-0.024	-0.032	-0.0032	-0.009	1	0.029	0.023	-0.06	0.02
	0.073	-0.016	0.0018	0.0074	0.015	0.06	0.016	0.051	0.029	1	-0.015	0.013	0.012
	-0.031	-0.068	-0.025	0.039	0.0061	0.021	0.052	0.023	0.023	-0.015	1	0.024	0.11
	0.036	-0.061	-0.0099	0.015	0.003	0.026	0.038	-0.029	-0.06	0.013	0.024	1	0.0035
	-0.0047	0.08	0.056	0.017	-0.039	0.011	-0.05	0.0065	0.02	0.012	0.11	0.0035	1
факторы, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при растяжении, МПа													
плотность при													



# Приложения. Репозиторий.

```
[ ] 1 # Запуск сервера Flask
2 if __name__ == "__main__":
3     app.run(debug=True, use_reloader=False)
```

```
* Serving Flask app "__main__" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
INFO:werkzeug: * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
INFO:werkzeug:127.0.0.1 - - [31/Mar/2023 17:01:34] "GET / HTTP/1.1" 200 -
```

Приложение выдает значение 2.914088010787964

Соотношение "матрица-наполнитель" для аналогичных показателей (позиция 3 датафрейма df\_clean) составляет 2.767918

```
[ ] 1 df_clean.iloc[3]
```

Соотношение матрица-наполнитель	2.767918
Плотность, кг/м3	2000.000000
модуль упругости, ГПа	748.000000
Количество отвердителя, м.%	111.860000
Содержание эпоксидных групп,%_2	22.267857
Температура вспышки, C_2	284.615385
Поверхностная плотность, г/м2	210.000000
Модуль упругости при растяжении, ГПа	70.000000
Прочность при растяжении, МПа	3000.000000
Потребление смолы, г/м2	220.000000
Угол нашивки	0.000000
Шаг нашивки	5.000000
Плотность нашивки	60.000000
Name: 3, dtype: float64	

Приложение прогнозирует соотношение "матрица-наполнитель"  
Введите "1" для прогноза, "2" для выхода  
1  
Введите данные для прогноза  
Плотность, кг/м3: 2000  
Модуль упругости, ГПа: 748  
Количество отвердителя, м.%.: 111.860000  
Содержание эпоксидных групп,%\_2: 22.267857  
Температура вспышки, C\_2: 284.615385  
Поверхностная плотность, г/м2: 210  
Модуль упругости при растяжении, ГПа: 70  
Прочность при растяжении, МПа: 3000  
Потребление смолы, г/м2: 220  
Угол нашивки: 0  
Шаг нашивки: 5  
Плотность нашивки: 60  
1/1 [=====] - 0s 63ms/step  
Прогнозное значение соотношения "матрица-наполнитель":  
2.914088  
Введите "1" для прогноза, "2" для выхода  
2

Приложение выдает значение 2.914088

Соотношение "матрица-наполнитель" для аналогичных показателей (позиция 3 датафрейма df\_clean) составляет 2.767918

```
[ ] 1 df_clean.iloc[3]
```

Соотношение матрица-наполнитель	2.767918
Плотность, кг/м3	2000.000000
модуль упругости, ГПа	748.000000
Количество отвердителя, м.%	111.860000
Содержание эпоксидных групп,%_2	22.267857
Температура вспышки, C_2	284.615385
Поверхностная плотность, г/м2	210.000000
Модуль упругости при растяжении, ГПа	70.000000
Прочность при растяжении, МПа	3000.000000
Потребление смолы, г/м2	220.000000
Угол нашивки	0.000000
Шаг нашивки	5.000000
Плотность нашивки	60.000000
Name: 3, dtype: float64	

wwwmyroot / DS\_graduate Public

<> Code Issues Pull requests Actions Projects

main 2 branches 0 tags

wwwmyroot Поправлена нумерация стра

dataset	GoogleCollab error при чтении .xlsx	2 hours ago
ipynb	Error with big file in github-repo. Grand renew.	1 hour ago
ipynb_raw/arx	Error with big file in github-repo. Grand renew.	1 hour ago
p_zapiska	Поправлена нумерация страниц в пояснительной записке.	30 minutes ago
presentation	Error with big file in github-repo. Grand renew.	1 hour ago
README.md	Error with big file in github-repo. Grand renew.	1 hour ago
TODO.org	Поправлены статусы.	last week
tree_branch-main.txt	Контроль дерева папок по веткам.	last week



## [ Выводы. ]

В ходе работы с использованием разработанных алгоритмов была проведена обработка экспериментальных данных модуля упругости при растяжении, прочности при растяжении и соотношения «матрица-наполнитель» с использованием языка программирования python.

Как показал анализ исходных данных, **корреляционная зависимость** между характеристиками композитов **крайне слабая и стремится к нулю**. Этот факт непосредственно повлиял на результат работы регрессионных моделей. **Все использованные модели показали низкую прогнозирующую способность**. Лучшим алгоритмом для прогноза модуля упругости при растяжении выбран AdaBoostRegressor, для прогнозирования прочности при растяжении – XGBRegressor.

Созданная для рекомендации соотношения «матрица-наполнитель» **нейронная сеть также плохо справилась с поставленной задачей прогноза**. Такие низкие показатели работы алгоритмов машинного обучения говорят о том, что прогнозирование свойств композиционных материалов – достаточно сложный процесс, требующий как знаний в области композиционных материалов, так и опыта в построении и использовании алгоритмов машинного обучения.

Полученный неудовлетворительный результат может также свидетельствовать о недостатках и ошибках в наборе исходных данных, недостаточно глубокой и детальной обработке данных, неточностях в выборе алгоритмов машинного обучения и их параметров.

Таким образом, для успешного решения задачи, поставленной в выпускной квалификационной работе, **необходимы более глубокие знания в области материаловедения и технологии конструкционных материалов, математического анализа и статистики, а также в области решения задач машинного обучения и обработки данных**. Более детальное изучение данных вопросов и консультация квалифицированных специалистов из указанных областей определенно положительно повлияют на уточнение подходов и оптимизацию алгоритмов для решения задачи прогнозирования конечных свойств композиционных материалов.





ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана



[do.bmstu.ru](https://do.bmstu.ru)