



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА по курсу «Data Science»

Прогнозирование конечных свойств новых  
материалов (композиционных материалов)

Соколовский Болеслав Леонидович



## Содержание работы

**1** Постановка задачи и описание используемых методов

**2** Разведочный анализ данных, Предобработка данных

**3** Разработка и обучение моделей. Тестирование моделей

**4** Нейронная сеть

**5** Разработка приложения, Создание удаленного репозитория

**Прогнозирование  
конечных свойств  
новых материалов  
(композиционных  
материалов).**



## Постановка задачи и описание используемых методов

**Цель исследовательской работы – разработать модели для прогноза модуля упругости при растяжении, прочности при растяжении и соотношения «матрица-наполнитель».**

Так как поставленные задачи прогнозирования параметров являются задачей регрессии, для их решения были выбраны следующие методы машинного обучения с учителем:

1. Линейная регрессия
2. Полиномиальная регрессия;
3. KNeighborsRegressor
4. RandomForestRegressor
5. GradientBoostingRegressor

Для прогноза рекомендованного значения соотношения «матрица-наполнитель»:

- Полносвязная нейронная сеть (2 варианта)



# Разведочный анализ данных, Предобработка данных

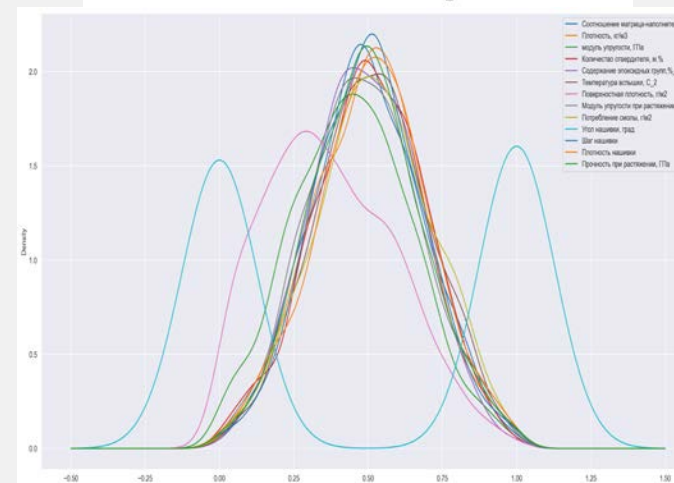
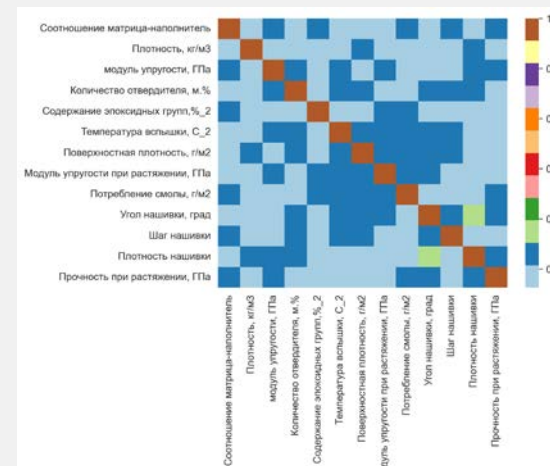
## Разведочный анализ данных

Датсеты df\_br и df\_nur были объединены по индексу, тип объединения INNER. Для формирования описания данных используя библиотеку ydata-profiling:

- пропусков и дубликатов в данных не выявлено.
- распределение величин, близкое к нормальному
- наблюдается отсутствие значимых взаимосвязей между переменными
- выявлены выбросы по всем признакам

## Предобработка

- удаление выявленных выбросов методом на основе межквартильного размаха(скошенное распределение 2 признаков)
- нормализацию данных с помощью MinMaxScaler()





## Разработка и обучение моделей

- При построении моделей был осуществлен поиск гиперпараметров модели с помощью поиска по сетке с перекрестной проверкой, количество блоков равно 10.
- При построении модели в соответствии с поставленной задачей 30% данных было оставлено на тестирование модели, на остальных 70% происходило обучение моделей.
- По итогам оценки лучшие результаты почти по всем выбранным метрикам для прогноза модуля упругости при растяжении показал метод RandomForestRegressor.
- Лучшие результаты для прочности при растяжении показал метод линейной регрессии. Хотя и параметры  $R^2$  в большинстве случаев отрицательны – что говорит о неудовлетворительной описательной способности моделей.



## ИТОГИ ПО МОДЕЛЯМ

model\_loss

|< < 10 rows > >| 10 rows x 5 columns

÷ целевая переменная ÷	÷ модель ÷	MAE ÷	MSE ÷	R2 ÷
0 Модуль упругости при растяжении	Linear Regression	0.145773	0.032470	-0.003036
1 Модуль упругости при растяжении	Polynomial Regression	0.156588	0.037613	-0.161910
2 Модуль упругости при растяжении	KNeighborsRegressor	0.148200	0.032793	-0.013020
3 Модуль упругости при растяжении	RandomForestRegressor	0.146204	0.032330	0.001306
4 Модуль упругости при растяжении	GradientBoostingRegressor	0.146091	0.032810	-0.013526
5 Прочность при растяжении	Linear Regression	0.141190	0.033028	-0.032068
6 Прочность при растяжении	Polynomial Regression	0.149222	0.036791	-0.149642
7 Прочность при растяжении	KNeighborsRegressor	0.141451	0.033193	-0.037208
8 Прочность при растяжении	RandomForestRegressor	0.142092	0.033202	-0.037486
9 Прочность при растяжении	GradientBoostingRegressor	0.140149	0.032722	-0.022485





# Нейронная сеть

## Первая модель

количество скрытых слоев – 2

количество нейронов на входном слое – 12

количество нейронов на 1 и 2 скрытых слоях – 25

количество нейронов на выходном слое – 1

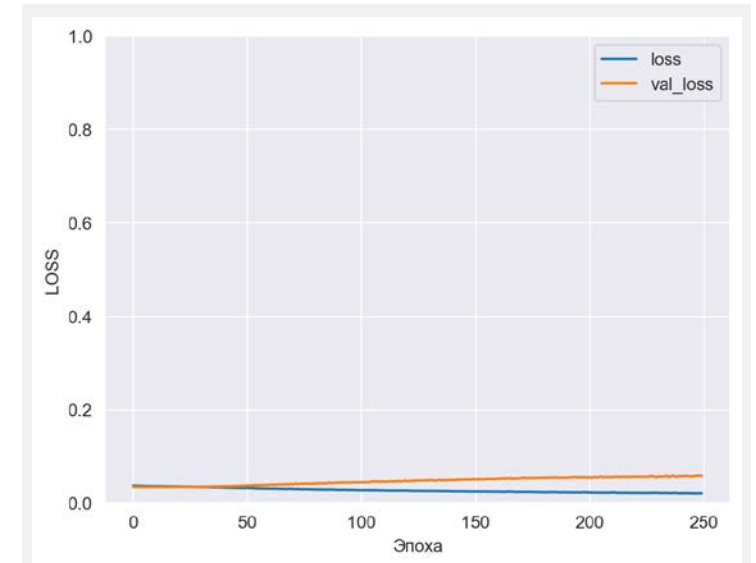
активационная функция «Relu», на выходном слое  
«Sigmoid»

оптимизатор «Adam»

функция потерь – MSE

метрика – MAE

Обучение модели происходило за 250 эпох.





## Вторая модель

количество скрытых слоев – 4

количество нейронов на входном слое – 120

количество нейронов на 1 и 2 скрытых слоях – 1440

количество нейронов на 3 скрытом слое – 60

количество нейронов на 4 скрытом слое – 20

количество нейронов на выходном слое – 1;

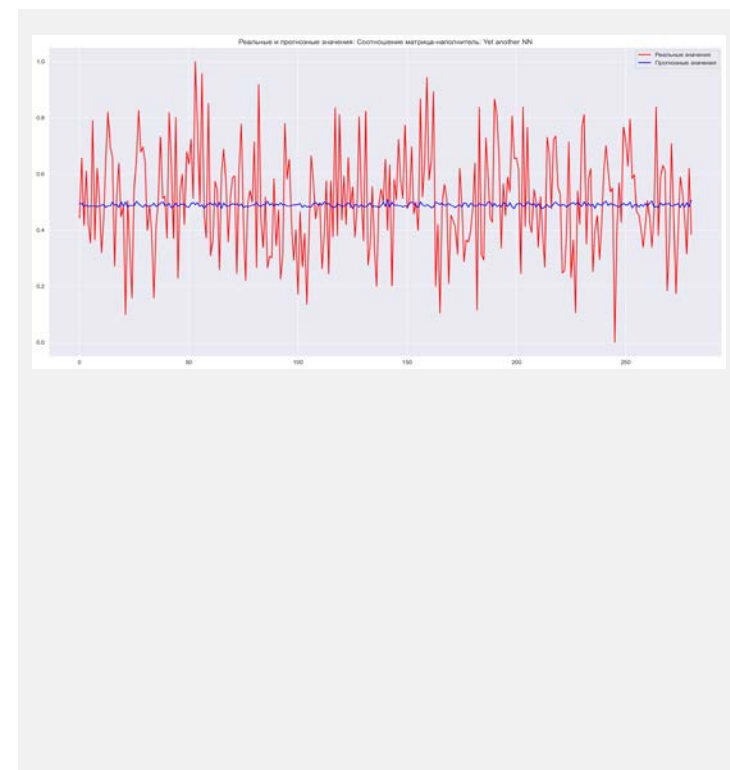
активационная функция «Relu», на выходном слое «Sigmoid»

оптимизатор «rmsprop»

функция потерь – MAE

метрика – MSE

В модели Yet\_another\_NN использовались точки фиксации минимального значения ошибки на тестовой выборке – чекпоинты. Согласно истории обучения, оптимальные параметры получены на 8 эпохе обучения. (MAE = 0.14713). В дальнейшем сохраним эту модель для использования в приложении. Обучение модели происходило также за 250 эпох.







# Нейронная сеть

12 rows × 5 columns					
целевая переменная	модель	MAE	MSE	R2	
0 Модуль упругости при растяжении	Linear Regression	0.145773	0.032470	-0.003036	
1 Модуль упругости при растяжении	Polynomial Regression	0.156588	0.037613	-0.161910	
2 Модуль упругости при растяжении	KNeighborsRegressor	0.148200	0.032793	-0.013020	
3 Модуль упругости при растяжении	RandomForestRegressor	0.146204	0.032330	0.001306	
4 Модуль упругости при растяжении	GradientBoostingRegressor	0.146091	0.032810	-0.013526	
5 Прочность при растяжении	Linear Regression	0.141190	0.033028	-0.032068	
6 Прочность при растяжении	Polynomial Regression	0.149222	0.036791	-0.149642	
7 Прочность при растяжении	KNeighborsRegressor	0.141451	0.033193	-0.037208	
8 Прочность при растяжении	RandomForestRegressor	0.142092	0.033202	-0.037486	
9 Прочность при растяжении	GradientBoostingRegressor	0.140149	0.032722	-0.022485	
10 Соотношение матрица-наполнитель	Sequential NN	0.169243	0.046851	-0.385044	
11 Соотношение матрица-наполнитель	Yet another NN	0.149472	0.034665	-0.024772	



ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ  
МГТУ им. Н.Э. Баумана

# Разработка приложения, Создание удаленного репозитория

## Приложение

Приложение было разработано с помощью Flask, HTML. Используется сохраненная модель нейронной сети Keras.

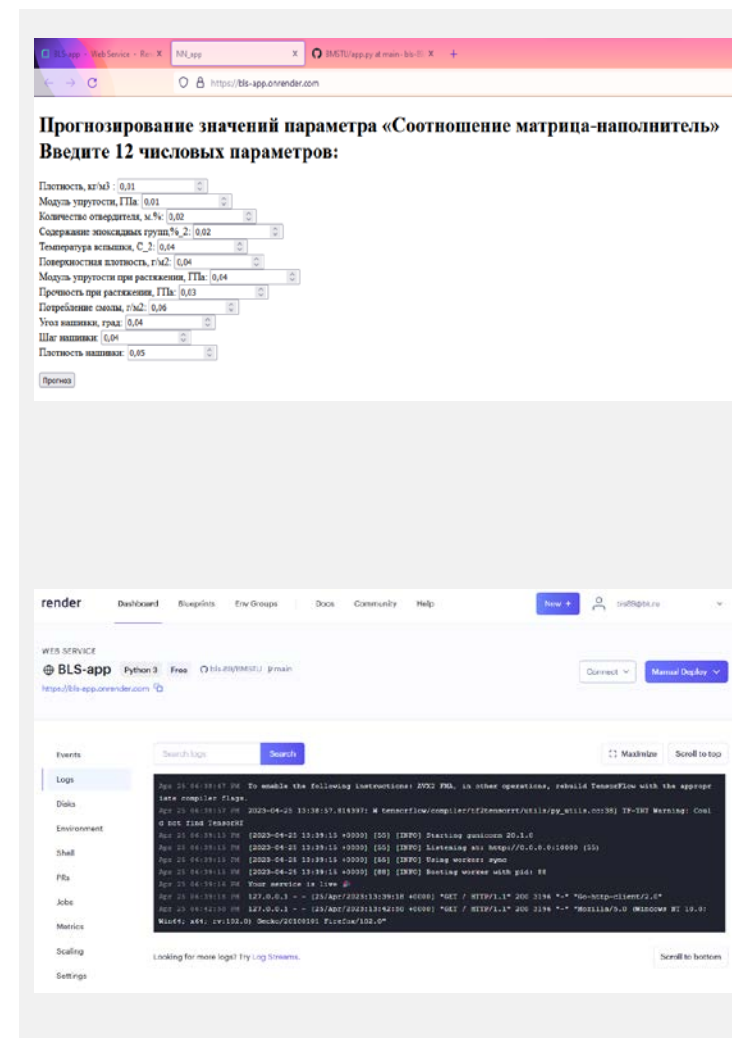
## Создание удаленного репозитория

Страница GitHub: [www.github.com/bls-89](https://www.github.com/bls-89)

Созданный репозиторий: [www.github.com/bls-89/BMSTU](https://www.github.com/bls-89/BMSTU)

Ссылка на приложение на render.com: [www.bls-app.onrender.com/](https://www.bls-app.onrender.com/)

Ссылка на страницу Kaggle.com: [www.kaggle.com/boleslav89](https://www.kaggle.com/boleslav89)





ЦЕНТР  
ДОПОЛНИТЕЛЬНОГО  
ОБРАЗОВАНИЯ

МГТУ им. Н.Э. Баумана



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