МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Московский государственный технический университет имени Н.Э. Баумана

(национальный исследовательский университет)»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**по курсу**

«Data Science»

**Тема:** «**Прогнозирование конечных свойств новых материалов**

**(композиционных материалов)»**

Слушатель Кошурников Дмитрий Сергеевич

Москва, 2022

# **Содержание**

[Содержание 2](#_Toc118829504)

[Введение 3](#_Toc118829505)

[1. Аналитическая часть 4](#_Toc118829506)

[1.1. Постановка задачи 4](#_Toc118829507)

[1.2. Описание используемых методов 6](#_Toc118829508)

[1.3. Разведочный анализ данных 9](#_Toc118829509)

[2. Практическая часть 16](#_Toc118829510)

[2.1. Предобработка данных 16](#_Toc118829511)

[2.2. Разработка и обучение модели 17](#_Toc118829512)

[2.3. Тестирование модели 18](#_Toc118829513)

[2.4. Разработка нейронной сети для 24](#_Toc118829514)

[2.5. Создание удалённого репозитория и загрузка 26](#_Toc118829515)

[2.6. Заключение 26](#_Toc118829516)

[2.7. Список использованной литературы и веб ресурсов. 27](#_Toc118829517)

# **Введение**

Композиционные материалы - это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними. Композиты обладают теми свойствами, которые не наблюдаются у компонентов по отдельности. При этом композиты являются монолитным материалом, т.е. компоненты материала неотделимы друг от друга без разрушения конструкции в целом. Яркий пример композита - железобетон. Бетон прекрасно сопротивляется сжатию, но плохо растяжению. Стальная арматура внутри бетона компенсирует его неспособность сопротивляться сжатию, формируя тем самым новые, уникальные свойства. Современные композиты изготавливаются из других материалов: полимеры, керамика, стеклянные и углеродные волокна, но данный принцип сохраняется. У такого подхода есть и недостаток: даже если мы знаем характеристики исходных компонентов, определить характеристики композита, состоящего из этих компонентов, достаточно проблематично. Для решения этой проблемы есть два пути: физические испытания образцов материалов, или прогнозирование характеристик. Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов (связующего и армирующего компонента).

На входе имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов. Кейс основан на реальных производственных задачах Центра НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана).

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

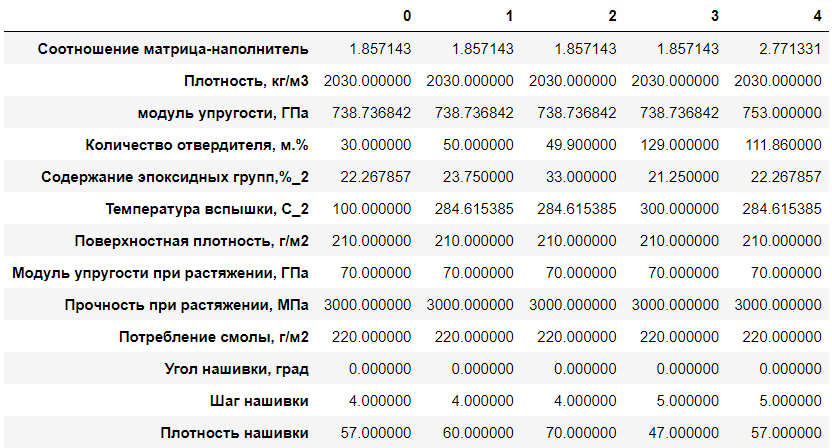
**Цель работы** – с помощью методов машинного обучения разработать модели для прогноза модуля упругости при растяжении, прочности при растяжении в зависимости от значений параметров композиционного материала.

В процессе исследовательской работы были разработаны и обучены несколько моделей, способные прогнозировать модуль упругости при растяжении и прочность при растяжении. Также была создана нейронная сеть, которая предлагает оптимальное значение параметра «матрицы - наполнитель». На основе нейронной сети было создано пользовательское веб - приложение на фреймворке Flask.

1. **Аналитическая часть**
   1. **Постановка задачи**

Для выпускной квалификационной работы были представлены 2 файла с данными в формате Excel: X\_bp.xlsx и X\_nup.xlsx. В первом файле содержались данные о 10 параметрах композиционного материала для 1024 образца (таблица из 10 столбцов и 1024 строк). Во втором файле содержались данные о 3 параметрах для 1041 образца (таблица из 3 столбцов и 1041 строки). Данные были объединены по индексу (тип объединения INNER) в один датасет, избыточные 17 строк с данными во втором файле были удалены (таблица 1). Объединённый датасет был сохранен в файл General.xlsx. Данные из этого файла использовались для дальнейшего разведочного анализа, поиска и удаления пропущенных значений, дублирующих значений, выбросов, а также для нормализации и стандартизации данных. Предобработанные данные затем использовались для построения моделей машинного обучения.

Таблица 1 - Результат объединения данных из двух файлов в один датасет (фрагмент)



Необходимо получить значения описательных статистик для каждого параметра, проверить наличие выбросов, дублирующих значений. Провести нормализацию и стандартизацию данных. Обучить несколько моделей машинного обучения для прогнозирования модуля упругости при растяжении и прочности при растяжении, сравнить результаты, получаемые с помощью этих моделей. Разработать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель для получения оптимальных эксплуатационных свойств композита. Создать веб-приложение с графическим интерфейсом, которое будет выдавать прогноз соотношения «матрица-наполнитель». Создать репозиторий в GitHub и разместить в нём написанный код на языке Python. Оформить файл README.

* 1. **Описание используемых методов**

Для обучения были выбраны следующие методы машинного обучения:

* метод опорных векторов;
* случайный лес;
* линейная регрессия;
* К-ближайших соседей;

**Метод опорных векторов (Support Vector Machines** — **SVM**) — это набор контролируемых методов обучения, используемых для классификации, регрессии и обнаружения выбросов.

*Преимущества:*

* Эффективен в пространствах больших размеров.
* Эффективен в случаях, когда количество измерений превышает количество образцов.
* Использует подмножество обучающих точек в функции принятия решений (называемых опорными векторами), поэтому это также эффективно с точки зрения памяти.
* Универсальность: для функции принятия решения могут быть указаны различны функции ядра. Предоставляются общие ядра, но также можно указать собственные ядра.

*Недостатки:*

* SVM не предоставляют напрямую оценки вероятностей, они рассчитываются с использованием дорогостоящей пятикратной перекрестной проверки.

**Метод случайного леса** (**Random Forest Regression**) — универсальный алгоритм машинного обучения, суть которого состоит в использовании ансамбля решающих деревьев. Само по себе решающее дерево предоставляет крайне невысокое качество классификации, но из-за большого их количества результат значительно улучшается. Также это один из немногих алгоритмов, который можно использовать в абсолютном большинстве задач. Благодаря своей гибкости Random Forest применяется для решения практически любых проблем в области машинного обучения. Сюда относятся задачи классификации (RandomForestClassifier) и регрессии (RandomForestRegressor).

*Преимущества:*

* Имеет высокую точность предсказания, которая сравнима с результатами градиентного бустинга.
* Не требует тщательной настройки параметров.
* Не чувствителен к масштабированию и к другим монотонным преобразованиям значений признаков.
* Редко переобучается. На практике добавление деревьев только улучшает композицию.

*Недостатки:*

* Для реализации алгоритма случайного дерева требуется значительный объем вычислительных ресурсов.
* Большой размер моделей.
* Построение случайного леса отнимает больше времени, чем деревья решений или линейные алгоритмы.
* В отличие от более простых алгоритмов, результаты случайного леса сложнее интерпретировать.

**Множественная линейная регрессия (MLR)** – это статистический метод, который использует несколько независимых переменных для прогнозирования результата переменной ответа. Цель множественной линейной регрессии (MLR) – смоделировать линейную связь между независимыми (независимыми) переменными и ответной (зависимой) переменной.

*Преимущества:*

* Простота реализации и моделирования.
* Понятность модели.

*Недостатки:*

* Линейная регрессия ограничена линейными отношениями. По своей природе линейная регрессия рассматривает только линейные отношения между зависимыми и независимыми переменными. То есть предполагается, что между ними существует прямая связь.
* Линейная регрессия рассматривает только среднее значение зависимой переменной: Линейная регрессия смотрит на соотношение между средним значением зависимой переменной и независимыми переменными.
* Линейная регрессия чувствительна к выбросам: выбросы - это удивительные данные. Выбросы могут быть одномерными (на основе одной переменной) или многомерными.
* Данные должны быть независимыми: линейная регрессия предполагает независимость данных. Это означает, что оценки одного предмета не имеют никакого отношения к оценкам другого. Это создает жесткое ограничение в приложениях кластеризации, где переменные должны быть кластеризованы на основе пространства и времени.

**Метод К-ближайших соседей** (**К-Nearest Neighbors**) – это простой алгоритм машинного обучения с учителем, который можно использовать для решения задач классификации и регрессии.

*Преимущества:*

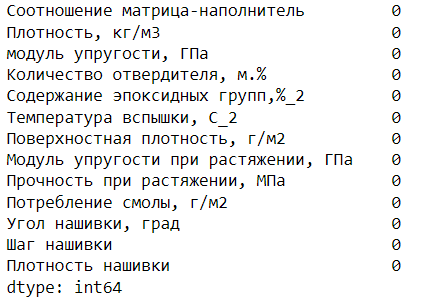
* Алгоритм прост и легко реализуем.
* Не чувствителен к выбросам.
* Нет необходимости строить модель, настраивать несколько параметров или делать дополнительные допущения.
* Алгоритм универсален. Его можно использовать для обоих типов задач: классификации и регрессии.

*Недостатки:*

* Алгоритм работает значительно медленнее при увеличении объема выборки, предикторов или независимых переменных.
* Из аргумента выше следуют большие вычислительные затраты во время выполнения.
* Всегда нужно определять оптимальное значение **k**.
  1. **Разведочный анализ данных**

Был проведён разведочный анализ данных и визуализация данных. В ходе разведочного анализа дублирующие значения (образцы) не были выявлены. Пропуски в данных также не были обнаружены (таблица 2).

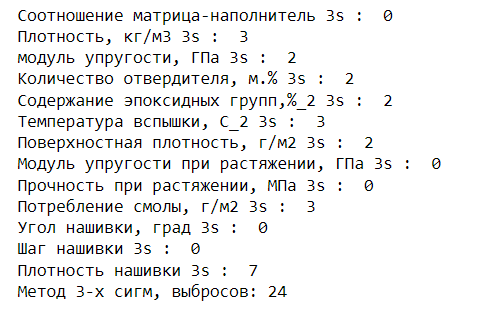
Таблица 2 - Проверка датасета на наличие пропущенных значений



Приступив к поиску выбросов, были отмечены некоторые важные особенности полученного датасета. Представленные для анализа данные характеризовались большим разбросом значений по каждому параметру (признаку, фактору), что делало выборку очень неоднородной. Общее количество параметров (признаков, факторов), по которым были распределены данные, составляло 13, из них три параметра рассматривались в качестве конечных эксплуатационных характеристик: модуль упругости, модуль упругости при растяжении и прочность при растяжении. Остальные параметры считались произвольно задаваемыми технологически при производстве композиционного материала. Высокие и низкие значения этих параметров не рассматривались в качестве выбросов. Значения трёх целевых параметров, характеризующих конечные эксплуатационные свойства композиционного материала (модуль упругости, модуль упругости при растяжении и прочность при растяжении), лежащие в областях низких и высоких значений, условно считались выбросами, и такие образцы были удалены из датасета.

Поиск и удаление выбросов является важной задачей для последующего использования данных для построения моделей машинного обучения и адекватного обучения моделей, что обеспечивает высокую точность предсказаний значений интересующих признаков. При этом удаление из выборки значений, лежащих в области высоких или низких областей, может представлять опасность в смысле искусственного искажения исходных данных. Такие значения могут указывать на наличие дополнительных групп данных, также принадлежащих к генеральной совокупности. Поэтому поиск и удаление выбросов должны проводиться осмысленно и очень осторожно, чтобы не нанести вред имеющимся данным и не исказить информацию, содержащуюся в данных, для достижения произвольных целей, которые преследует дата-сайентист, исследующий эти данные.

Таблица 3 - Количество выбросов для каждого из параметров датасета, выявленное методом 3ϭ



В качестве базового метода выявления выбросов был выбран метод трёх сигм (3ϭ). Всего по всем признакам было выявлено 24 значения, превышающие отклонение в 3ϭ и 2 значения по одному из целевых параметров – модулю упругости (таблица 3). Было признано, что результатов этого метода недостаточно, поэтому также учитывались данные диаграмм типа «ящик с усами» с межквартильными расстояниями для трёх целевых параметров (рисунок 1). Значения, превышающие отклонение в 3ϭ и выходящие за пределы межквартильного размаха 25-75, были удалены. Конечной размер датасета составил 999 строки с данными. Описательная статистика конечного датасета без выбросов приведена в таблице 4.

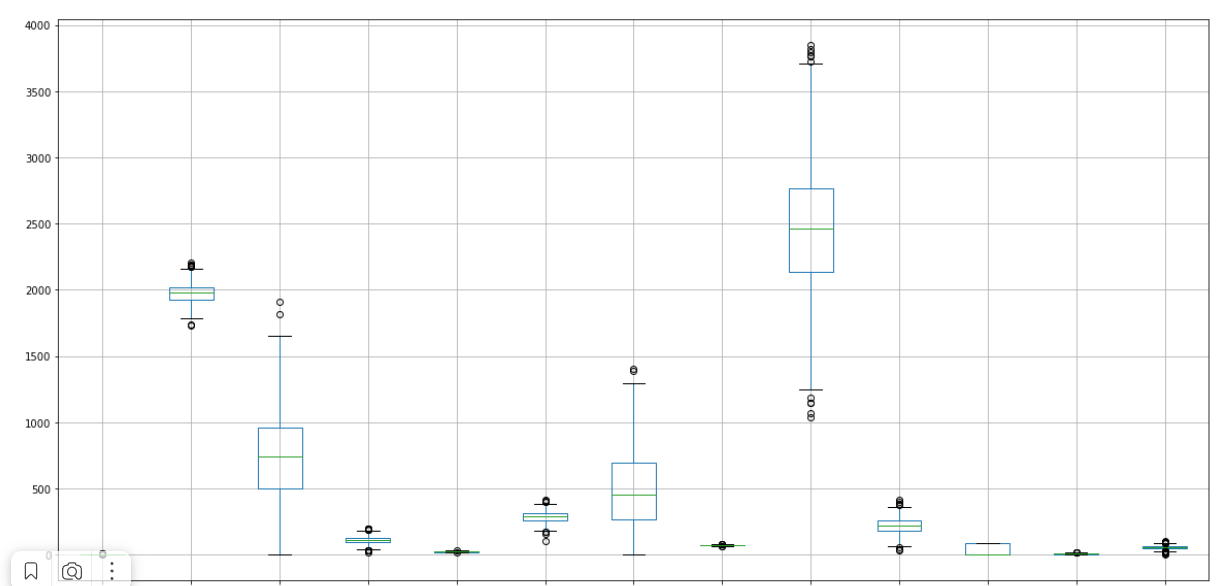
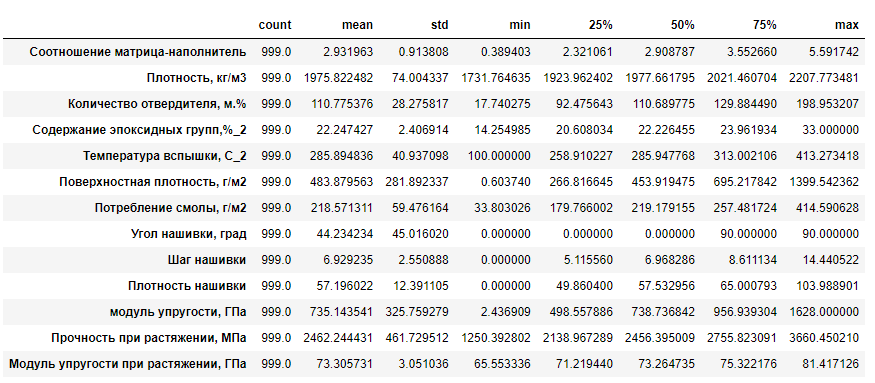


Рисунок 1 - Диаграмма типа «ящик с усами» для значений всех признаков

Таблица 4 - Описательная статистика датасета после удаления выбросов



На следующем этапе проводилось изучение распределение значений каждого из параметров датасета (рисунок 2). Как видно из графиков, по характеру распределения большая часть параметров (кроме параметров угол нашивки и поверхностная плотность) имела приближённо нормальное распределение.

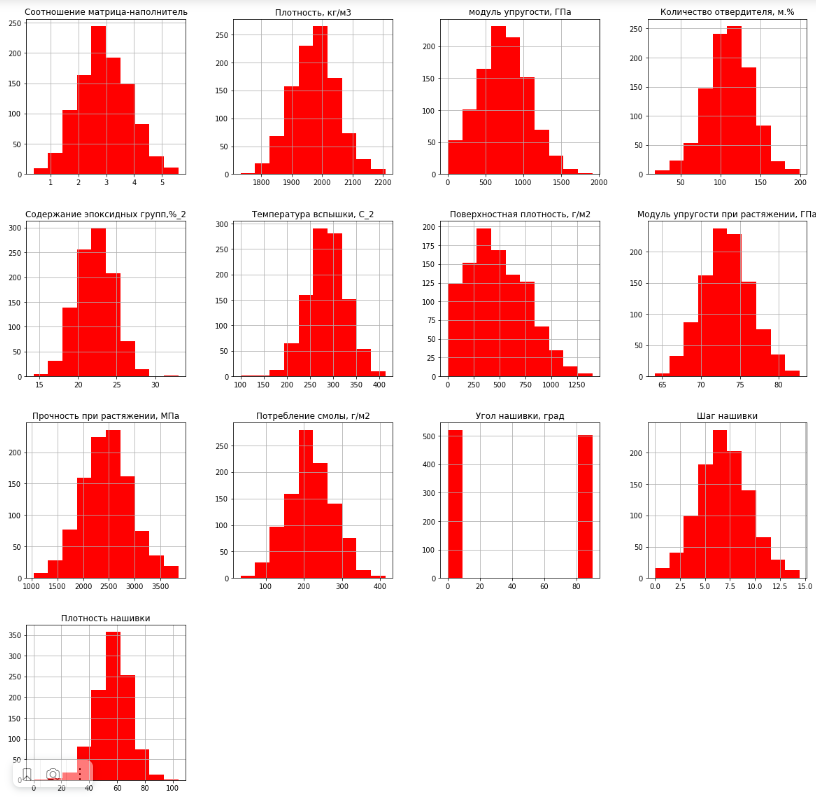


Рисунок 2 – Гистограммы распределения каждого параметра

Далее проводилось изучение зависимостей между всеми параметрами путём построения графиков рассеяния точек попарно для всех пар параметров (всего 13 параметров).

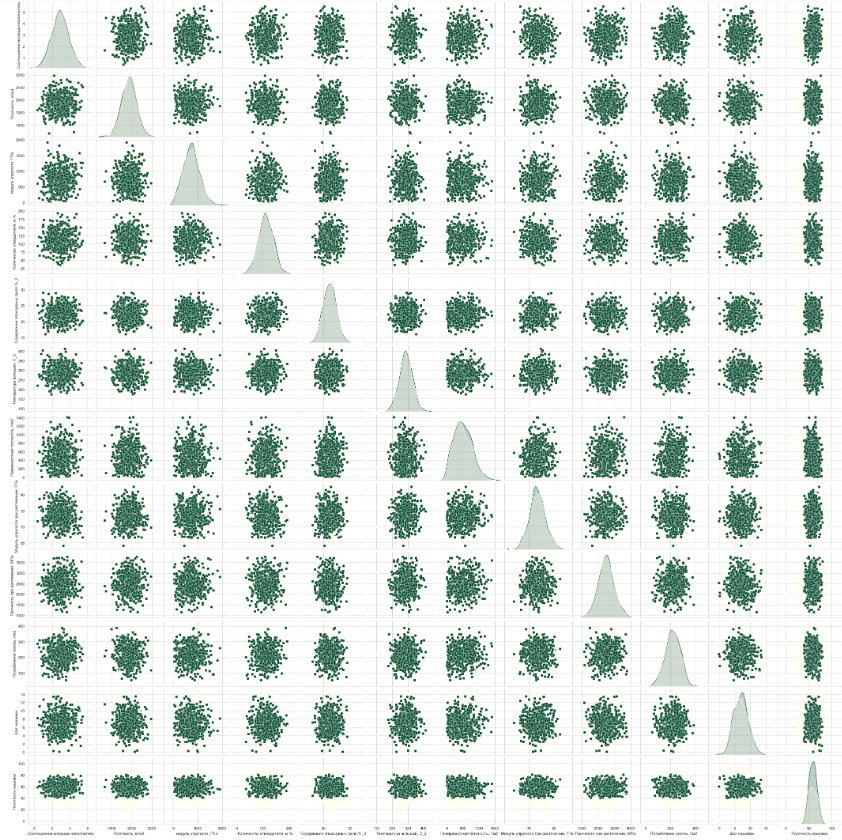


Рисунок 3 - Попарные графики рассеяния точек для всех параметров датасета

Как видно из графиков рассеяния точек для всех пар параметров чёткие зависимости между параметрами датасета не выявляются (рисунок 3).

Также оценивались корреляционные зависимости для всех пар параметров с помощью коэффициента корреляции Спирмена. Результаты корреляционного анализа в виде тепловой карты представлены на рисунке 4.

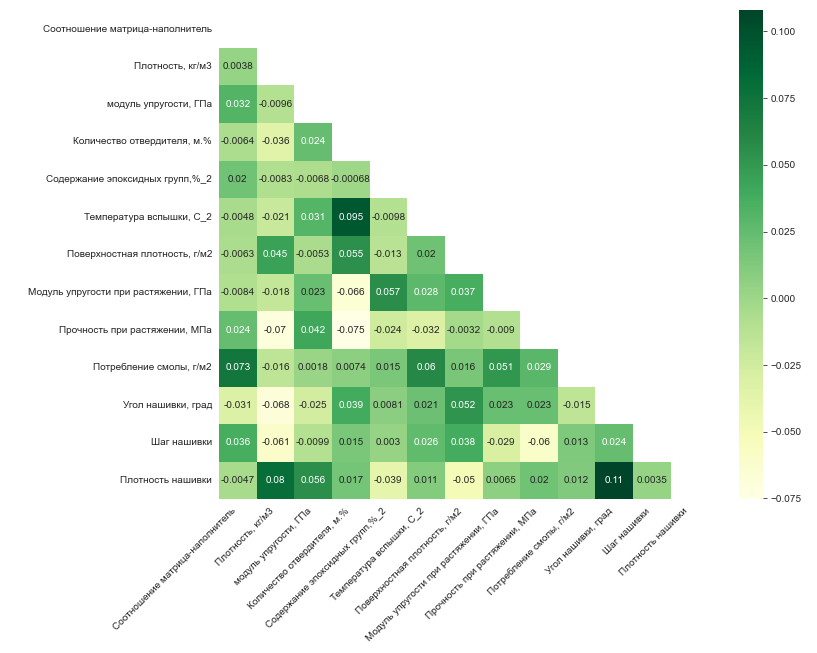


Рисунок 4 - Тепловая карта коэффициентов корреляции Спирмена между параметрами

Как видно из представленной тепловой карты, корреляционные зависимости между параметрами датасета практически отсутствуют (коэффициенты корреляции близки к нулю).

Для адекватной оценки полученных результатов разведочного анализа данных потребовалось ознакомиться с несколькими руководствами по материаловедению и композиционным материалам. При разведочном анализе данных не было установлено ни одной значимой корреляции ни для одной пары параметров. Особенно это странно для пары "соотношение матрица/наполнитель-плотность", для которой можно было ожидать наличие довольно сильной корреляции. Этот факт настораживает в отношении дальнейшей работы с полученными данными и возможности машинного обучения на этих данных. Единственное объяснение, которое можно предложить заключается в примерно одинаковой плотности матрицы и наполнителя, что представляется маловероятным. Также следует отметить, что нам ничего не известно о самом композите. По всей видимости, речь идёт о каком-то композите на основе полимерных эпоксидных смол.

1. **Практическая часть**
   1. **Предобработка данных**

В ходе подготовки данных для дальнейшего их использования (машинное обучение) применялись процедуры нормализации и стандартизации данных. Гистограммы распределения параметров после проведения нормализации приведены на рисунке 5.

Нормализация (масштабирование) и стандартизация данных могут иметь важное значение для адекватной настройки моделей машинного обучения. Нормализованные и стандартизированные данные сохранялись в отдельные файлы General\_norm.xlsx и General\_std.xlsx соответственно.

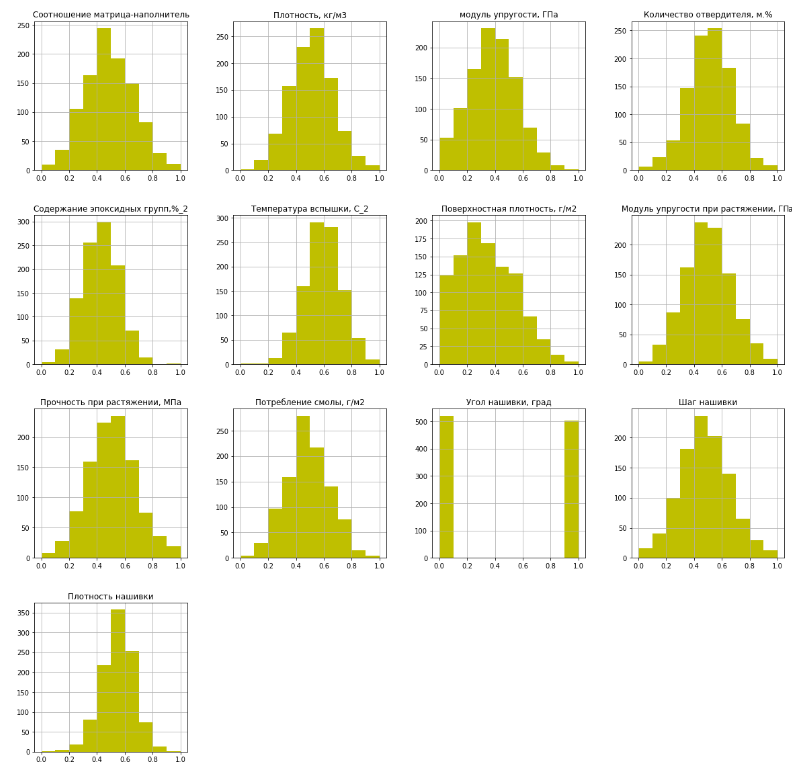


Рисунок 5 – Гистограммы распределения каждого параметра после нормализации

* 1. **Разработка и обучение модели**

Разработка и обучение моделей машинного обучения проводились для предсказания значений двух эксплуатационных (целевых) параметров прочность при растяжении и модуль упругости при растяжении на основании других 11 характеристик (признаков) композиционного материала. Применялись методы опорных векторов (Support Vector Regression), случайного леса (Random Forest Regression), множественной линейной регрессии (Multiple Linear Regression), К ближайших соседей (K Nearest Neighbors). Имеющиеся данные разделялись на обучающую выборку (70% данных) и тестовую выборку (30% данных), на которых проводилось обучение и тестирование выбранных моделей. Подбор оптимальных гиперпараметров осуществлялся с помощью процедуры GridSearchCV для каждой из выбранных моделей.

* 1. **Тестирование модели**

Проводилось обучение и тестирование модели машинного обучения методом опорных векторов (Support Vector Regression). Результаты измерений и прогнозируемые значения прочности при растяжении на тестовой выборке представлены на рисунке 6. Как видно из представленных графиков, результат обучения модели неудовлетворительный: были получены низкое значение коэффициента детерминации и высокое значение средней абсолютной ошибки на обучающей и тестовой выборках. Проводился поиск значений гиперпараметров с помощью метода GridSearchCV, найденные оптимальные значения гиперпараметров существенно не повысили точность работы модели.

Проводилось обучение модели методом случайного леса (Random Forest Regression). Результаты измерений и прогнозируемые значения прочности при растяжении на тестовой выборке представлены на рисунке 6. Как видно из представленных графиков, результат обучения модели неудовлетворительный: были получены низкое значение коэффициента детерминации и высокое значение средней абсолютной ошибки на обучающей и тестовой выборках. Проводился поиск значений гиперпараметров с помощью метода GridSearchCV, найденные

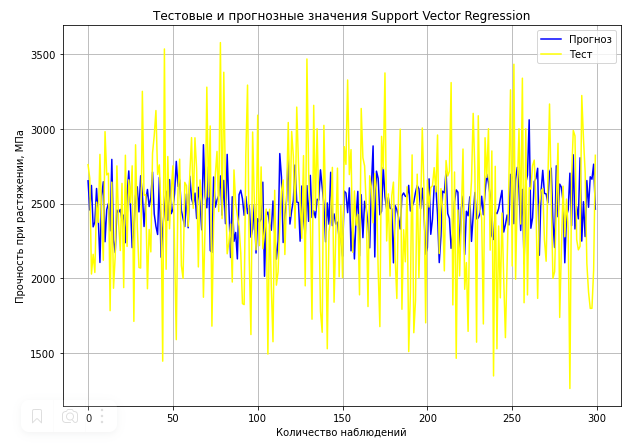


Рисунок 6 - Результат обучения с помощью метода опорных векторов (Support Vector Regression) для прогнозирования значений параметра прочность при растяжении.

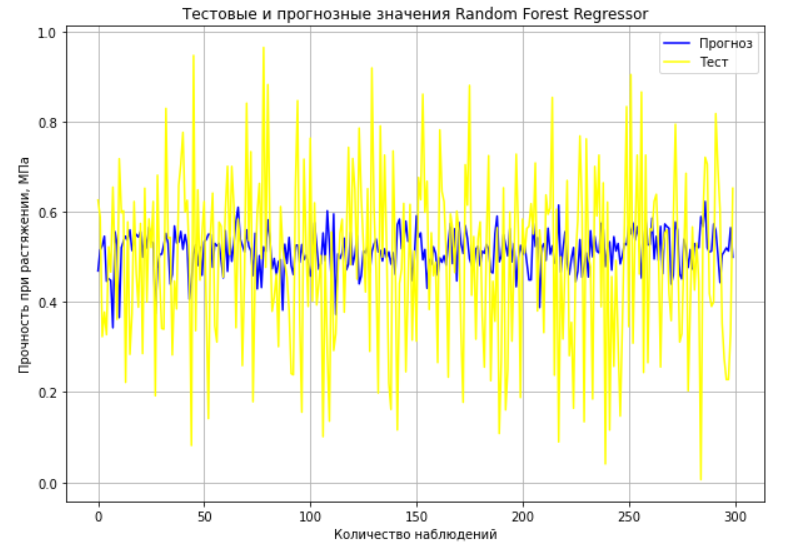


Рисунок 7 - Результат обучения с помощью метода случайного леса (Random Forest Regression) для прогнозирования значений параметра прочность при растяжении.

оптимальные значения гиперпараметров существенно не повысили точность работы модели.

Проводилось обучение модели методом множественной линейной регрессии (Multiple Linear Regression). Результаты измерений и прогнозируемые значения прочности при растяжении на тестовой выборке представлены на рисунке 6. Как видно из представленных графиков, результат обучения модели неудовлетворительный: были получены низкое значение коэффициента детерминации и высокое значение средней абсолютной ошибки на обучающей и тестовой выборках. Проводился поиск значений гиперпараметров с помощью метода GridSearchCV, найденные оптимальные значения гиперпараметров существенно не повысили точность работы модели.

Проводилось обучение модели методом К ближайших соседей (K Nearest Neighbors). Результаты измерений и прогнозируемые значения прочности при растяжении на тестовой выборке представлены на рисунке 6. Как видно из представленных графиков, результат обучения модели неудовлетворительный: были получены низкое значение коэффициента детерминации и высокое значение средней абсолютной ошибки на обучающей и тестовой выборках. Проводился поиск значений гиперпараметров с помощью метода GridSearchCV, найденные оптимальные значения гиперпараметров существенно не повысили точность работы модели.

Таким образом, были получены результаты работы четырёх моделей машинного обучения для предсказания значений прочности при растяжении для композиционного материала. Из полученных результатов видно, что точность работы моделей низкая, и выбранные модели не могут обеспечить надёжного предсказания прочности при растяжении для исследуемого композитного материала. Далее проводилось обучение и тестирование тех же выбранных моделей машинного обучения для прогнозирования значений параметра модуль упругости при растяжении для того же композита.

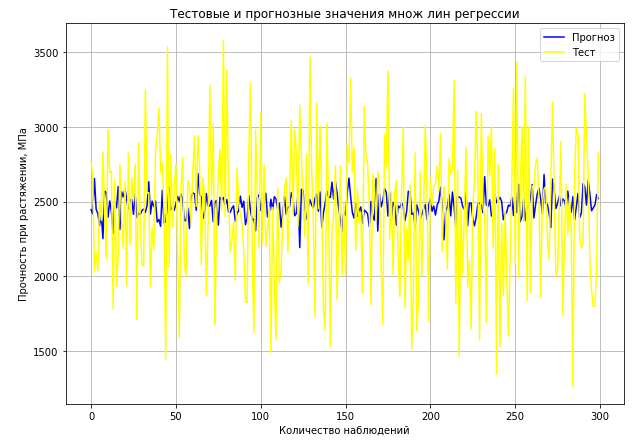


Рисунок 8 - Результат обучения с помощью метода множественной линейной регрессии (Multiple Linear Regression) для прогнозирования значений параметра прочность при растяжении.

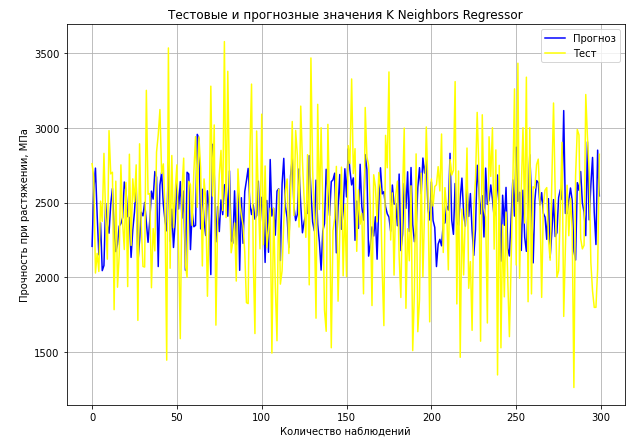
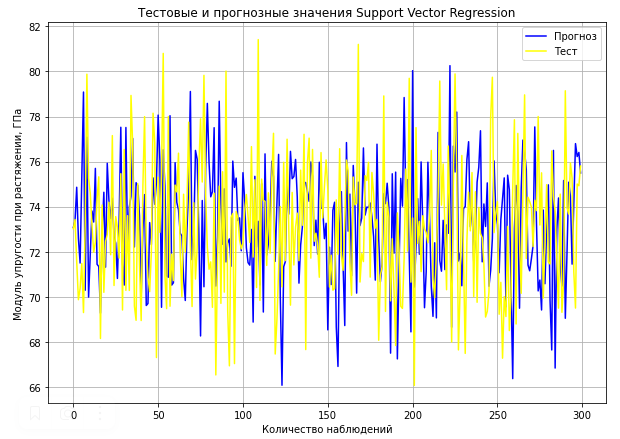


Рисунок 9 - Результат обучения с помощью метода К ближайших соседей (K Nearest Neighbors) для прогнозирования значений параметра прочность при растяжении.

Таблица 5 - Показатели точности предсказания значений прочности при растяжении на обучающей и тестовой выборках для всех выбранных моделей.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Показатель** | Support Vector Regression | Random Forest Regression | Multiple Linear Regression | K Nearest Neighbors |
| Коэффициент детерминации (обучающая выборка) | 0.39 | 0.44 | 0.03 | 0.23 |
| MAE (обучающая выборка) | 246 | 283 | 366 | 327 |
| Коэффициент детерминации (тестовая выборка) | -0.12 | -0.06 | -0.03 | -0.15 |
| MAE (тестовая выборка) | 389 | 375 | 373 | 392 |

Результаты работы выбранных моделей машинного обучения по предсказанию модуля упругости при растяжении в основном существенно не отличались от таковых для параметра прочность растяжении на обучающей и тестовой выборках. Результаты обучения моделей следует также считать неудовлетворительным: были получены низкие значения коэффициента детерминации и высокие значения средней абсолютной ошибки на обучающей и тестовой выборках. Проводился поиск значений гиперпараметров с помощью метода GridSearchCV, найденные оптимальные значения гиперпараметров существенно не повысили точность работы моделей. Исключение составила модель на основе метода опорных векторов: для неё были получены относительно высокое значение коэффициента детерминации на тестовой выборке и низкое значение средней абсолютной ошибки.

Рисунок 10 - Результат обучения с помощью метода опорных векторов (Support Vector Regression) для прогнозирования значений параметра модуль упругости при растяжении.

Результаты работы выбранных моделей машинного обучения по прогнозированию модуля упругости при растяжении представлены в таблице 6.

Данные неудовлетворительные результаты работы моделей машинного обучения могут свидетельствовать о некачественно собранных данных, либо об отсутствии зависимости конечных эксплуатационных характеристик от выбранных параметров композита. Возможно, существуют другие параметры, не вошедшие в данный анализ, оказывающие решающее влияние на эксплуатационные свойства композитного материала.

Таблица 6 - Показатели точности предсказания значений модуля упругости при растяжении на обучающей и тестовой выборках для всех выбранных моделей.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Показатель** | Support Vector Regression | Random Forest Regression | Multiple Linear Regression | K Nearest Neighbors |
| Коэффициент детерминации (обучающая выборка) | 0.91 | 0.42 | 0.01 | 0.16 |
| MAE (обучающая выборка) | 0.92 | 0.12 | 0.16 | 0.14 |
| Коэффициент детерминации (тестовая выборка) | -0.65 | -0.06 | -0.02 | -0.18 |
| MAE (тестовая выборка) | 3.07 | 0.15 | 0.15 | 0.16 |

* 1. **Разработка нейронной сети для**

**прогнозирования соотношения «матрица – наполнитель».**

Для решения задачи прогнозирования оптимального соотношения «матрица-наполнитель» была создана нейронная сеть по типу многослойного перцептрона. Оценивалась функция потерь и точность. Для достижения наилучших показателей работы нейронной сети подбирались основные параметры — количество слоёв, количество нейронов, функции активации слоёв.

Результаты оказались неудовлетворительными: с помощью разработанной нейронной сети не удалось добиться прогнозирования оптимального соотношения «матрица-наполнитель».

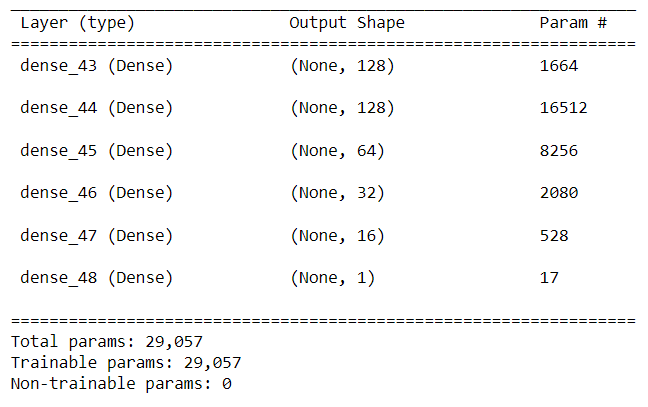


Рисунок 11 - Архитектура разработанной нейронной сети

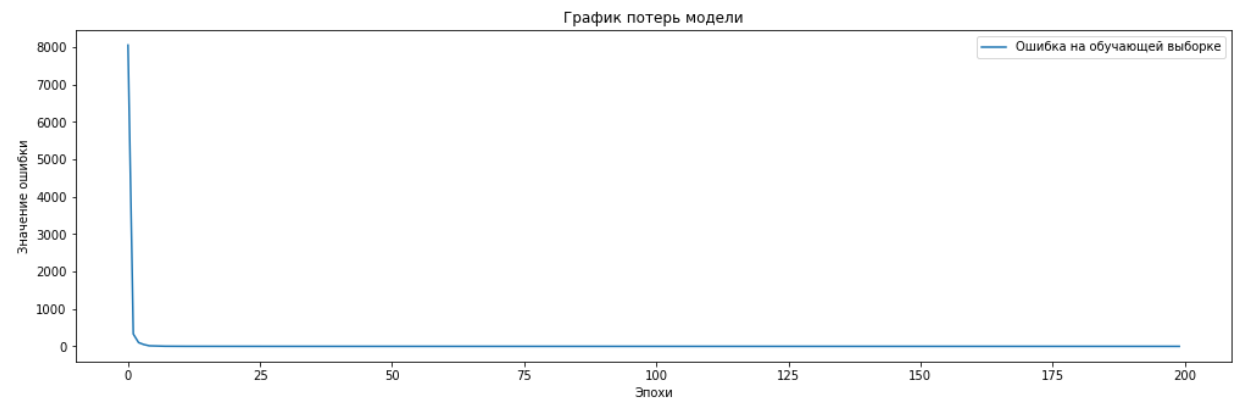


Рисунок 12 - График обучения нейронной сети

**Разработка приложения**

Было разработано пользовательское веб-приложение, которое выдаёт результат прогноза соотношения матрица-наполнитель при известных других параметрах композита. Приложение было создано с помощью библиотеки Flask и удобно для практического применения.

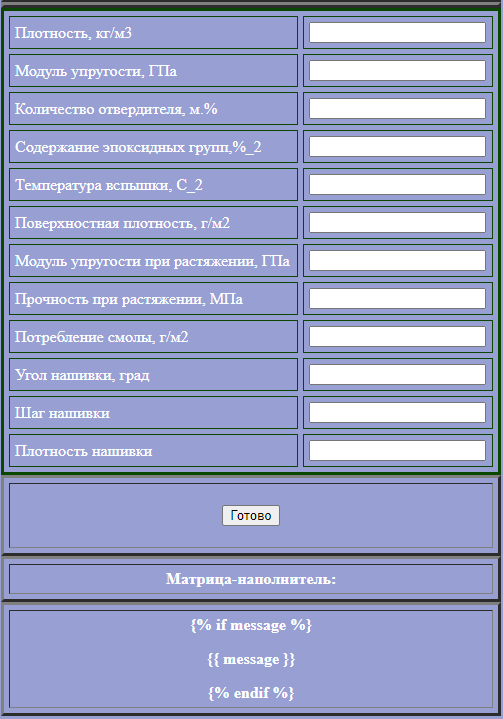


Рисунок 13 – графический интерфейс пользовательского приложения

* 1. **Создание удалённого репозитория и загрузка**

Создан репозиторий на github.com, загружены данные по исследованию.

Ссылка на репозиторий: https://github.com/Mitrich500/My-scripts

* 1. **Заключение**

В качестве рабочей гипотезы было предположено, что параметры композиционного материала (такие как соотношение матрица-наполнитель, температура вспышки, угол нашивки и др.) должны определять его конечные эксплуатационные свойства или, по меньшей мере, оказывать влияние на них. Однако, анализ представленных данных не выявил значимых закономерностей между характеристиками композиционного материала и его эксплуатационными свойствами. Данный факт, вероятнее всего, указывает на недостаточность представленных данных или низкое качество сбора данных. В качестве рекомендации можно предложить улучшить качество сбора данных, увеличить количество оцениваемых образцов и включить в анализ какие-либо недостающие параметры, которые могут иметь значение для конечных эксплуатационных свойств композиционного материала.

* 1. **Список использованной литературы и веб ресурсов.**

1. Берикашвили В.Ш. Статистическая обработка данных, планирование эксперимента и случайные процессы: учебное пособие для вузов/ Берикашвили В. Ш., Оськин С. П. - 2-е изд., испр. и доп. - М.: Юрайт, 2021. - 163 с.
2. Бизли Д. Python. Подробный справочник: учебное пособие. – Пер. с англ. – СПб.: Символ-Плюс, 2010. – 864 с.
3. Богодухов С. И., Козик Е. С. Материаловедение: учебник / С. И. Богодухов, Е. С. Козик. — Старый Оскол: ТНТ, 2016. — 536 с.
4. Гафаров, Ф.М., Галимянов А.Ф. Искусственные нейронные сети и приложения: учеб. пособие /Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Издательство Казанского университета, 2018. – 121 с.
5. Грас Д. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербурr, 2021. - 416 с.
6. Дауни Аллен Б. Основы Python. Научитесь думать как программист - М.: Манн, Иванов и Фербер, 2021. - 304 с.
7. Дауни Аллен Б. Рекомендательные системы на практике / пер. с англ. Д. М. Павлова. – М.: ДМК Пресс, 2020. – 448 с.
8. Иванов Д.А., Ситников А.И., Шляпин С.Д – Композиционные материалы: учебное пособие для вузов, 2019. - 233 с.
9. Комаров О.С. Материаловедение и технология конструкционных материалов: учебник / О.С. Комаров, В.Н. Ковалевский, Л.Ф. Керженцева и др.; под общ. ред. О.С. Комарова. — 3-еизд., испр. и доп. — Минск: Новое знание, 2009. — 671 с.
10. Кушнер В.С. Материаловедение: учеб. для студентов вузов /В. С. Кушнер, А. С. Верещака, А. Г. Схиртлаздзе, Д. А. Негров, О. Ю. Бургонова.; под ред. В. С. Кушнера. - Омск: Изд-во ОмГТУ, 2008. – 232 с.
11. Любанович Билл. Простой Python. Современный стиль программирования. — СПб.: Питер, 2016. — 480 с.
12. Орельен Ж. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. Пер. с англ. - СПб.: ООО «Альфа-книга», 2018. - 688 с.
13. Пал С., Джулли А. Библиотека Keras - инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / пер. с англ. А. А. Слинкин. - М.: ДМК Пресс, 2018. – 294 с.
14. Плас Дж. Вандер, Python для сложных задач: наука о данных и машинное обучение. Санкт-Петербург: Питер, 2018. - 576 с.
15. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.. : Пер. с англ. - М. : Издательский дом ‘‘Вильямс’’, 2007. - 1408 с.
16. Роббинс Д. HTML5: карманный справочник, 5-е издание.: Пер. с англ. - М.: ООО «И.Д. Вильямс», 2015. - 192 с.
17. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.
18. Скиена, Стивен С. Наука о данных: учебный курс.: Пер. с англ. - СПб.: ООО "Диалектика", 2020. - 544 с.
19. Харрисон М. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python. – СПб.: ООО «Диалектика», 2020. – 320 с.
20. Шитиков В. К., Мастицкий С. Э. (2017) Классификация, регрессия, алгоритмы Data Mining с использованием R. - Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining> . 351 с.
21. Ын Анналин, Су Кеннет. Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. - Спб.: Питер, 2019. - 208 с.
22. Nisbet R., Elder J., Miner G. Handbook of Statistical Analysis and Data Mining Applications. - Academic Press, 2009. — 864 p.
23. Документация по библиотеке keras: – Режим доступа: <https://keras.io/api/>.(дата обращения: 08.09.2022).
24. Документация по библиотеке matplotlib: – Режим доступа: <https://matplotlib.org/stable/users/index.html>. (дата обращения: 10.08.2022)
25. Документация по библиотеке numpy: – Режим доступа: [https://numpy.org/doc/1.22/user/index.html#user](_blank). (дата обращения: 03.08.2022).
26. Документация по библиотеке pandas: – Режим доступа: [https://pandas.pydata.org/docs/user\_guide/index.html#user-guide](_blank). (дата обращения: 04.08.2022).
27. Документация по библиотеке scikit-learn: – Режим доступа: <https://scikit-learn.org/stable/user_guide.html>. (дата обращения: 05.08.2022).
28. Документация по библиотеке seaborn: – Режим доступа: <https://seaborn.pydata.org/tutorial.html>. (дата обращения: 06.08.2022).
29. Документация по библиотеке Tensorflow: – Режим доступа: <https://www.tensorflow.org/overview> (дата обращения: 10.09.2022).
30. Документация по языку программирования python: – Режим доступа:  <https://docs.python.org/3.8/index.html>. (дата обращения: 28.07.2022).
31. Краткий обзор алгоритма машинного обучения Метод Опорных Векторов (SVM) – Режим доступа: <https://habr.com/ru/post/428503/> (дата обращения 07.09.2022).
32. Метод k-ближайших соседей (k-nearest neighbour): – Режим доступа: <https://proglib.io/p/metod-k-blizhayshih-sosedey-k-nearest-neighbour-2021-07-19> (дата обращения: 25.08.2022).
33. Пять алгоритмов регрессии в машинном обучении, о которых вам следует знать: – Режим доступа: https://habr.com/ru/company/vk/blog/513842/(дата обращения: 01.09.2022).
34. Руководство по быстрому старту в flask: – Режим доступа: <https://flask-russian-docs.readthedocs.io/ru/latest/quickstart.html>. (дата обращения: 09.10.2022).