

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

# **Методы искусственного интеллекта и машинное обучение**

Методические указания для проведения лабораторных работ  
Для студентов направления 09.03.04  
«Программная инженерия»

Автор-составитель В. В. Воронина

Ульяновск  
УлГТУ  
2022

УДК 004 (076)

ББК 32.973-018.1я7

В12

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук,  
доцент К.В. Святков.

Рекомендовано научно-методической комиссией факультета  
информационных систем и технологий в качестве практикума.

**Воронина, Валерия Вадимовна**

В12 Методы искусственного интеллекта и машинное обучение : практикум для проведения лабораторных занятий для студентов направления 09.03.04 «Программная инженерия» / В. В. Воронина. – Ульяновск : УлГТУ, 2022. – 35 с.

Составлен в соответствии с учебным планом направления 09.03.04. Цель данного практикума – ориентировать студентов на содержание и порядок выполнения лабораторных задач во время прохождения ими курсов «Методы искусственного интеллекта» и «Машинное обучение». Даются задания на лабораторные работы.

Работа подготовлена на кафедре «Информационные системы».

**УДК 004 (076)**

**ББК 32.973-018.1я7**

© В. В. Воронина, составление 2022.

© Оформление. УлГТУ, 2022.

## Оглавление

Введение.....	4
Массивы данных для лабораторных работ.....	4
Лабораторные работы по курсу «Методы искусственного интеллекта» .....	7
Лабораторная работа 1. Веб-сервис для загрузки и представления данных.....	9
Лабораторная работа 2. Веб-сервис для простейшего анализа данных .....	10
Лабораторная работа 3. Веб-сервис для визуализации данных .....	18
Лабораторная работа 4. Веб-сервис «Фильтр Блума» и интеграция сервисов.....	19
Лабораторная работа 5. Веб-сервис «Парная линейная регрессия» .....	19
Лабораторная работа 6. Веб-сервис «Дерево решений» .....	21
Лабораторная работа 7. Веб-сервис «Кластеризация» .....	22
Лабораторная работа 8. Генетический алгоритм .....	24
Лабораторная работа 9. Нечеткая логика .....	28
Лабораторная работа 10. Лингвистические переменные и шкалы .....	32
Лабораторная работа 11. Нечеткая кластеризация объектов.....	34
Лабораторная работа 12. Нечеткий логический вывод .....	37
Лабораторная работа 13. Анализ текстов .....	37
Лабораторная работа 14. Нейронные сети.....	39
Лабораторные работы по курсу «Машинное обучение» .....	40
Лабораторная работа 1. Работа с типовыми наборами данных и различными моделями .....	40
Лабораторная работа 2. Ранжирование признаков .....	44
Лабораторная работа 3. Деревья решений.....	47
Лабораторная работа 4. Кластеризация .....	48
Лабораторная работа 5. Регрессия.....	48
Лабораторная работа 6. Нейронная сеть.....	49
Лабораторная работа 7. Рекуррентная нейронная сеть и задача генерации текста .....	49
Заключение .....	<b>Ошибка! Закладка не определена.</b>
Список литературы .....	51

## Введение

Данные методические указания содержат задания и ссылки на теоретический материал, необходимый для прохождения курсов «Методы искусственного интеллекта» (два семестра) и «Машинное обучение» (один семестр).

В первом семестре курса «Методы искусственного интеллекта» студентам необходимо познакомиться с разработкой веб-сервисов на языке python для решения задачи анализа данных с использованием методов искусственного интеллекта и интеграцией разработок в единую систему. Во втором семестре студенты продолжают знакомиться с методами искусственного интеллекта, решая на языке python небольшие учебные задачи.

В курсе «Машинное обучение» студентам необходимо познакомиться с библиотечными реализациями моделей и методов искусственного интеллекта на языке Python, применив их для решения задачи анализа данных по варианту.

### Массивы данных для лабораторных работ

В таблице 1 представлены ссылки для скачивания наборов данных (файлы в формате csv), необходимые для выполнения лабораторных работ по курсам «Методы искусственного интеллекта» и «Машинное обучение». Если на момент выдачи задания какая-то ссылка будет являться не рабочей, то студенту необходимо обратиться к преподавателю для получения данных. Дата актуальности ссылок – 1.08.2022.

Таблица 1. Данные для анализа

№ варианта	Название набора данных	URL для скачивания
1.	Объекты вокруг Земли	<a href="https://www.kaggle.com/datasets/sameerpani/nasa-nearest-earth-objects">https://www.kaggle.com/datasets/sameerpani/nasa-nearest-earth-objects</a>

№ варианта	Название набора данных	URL для скачивания
2.	Набор данных для анализа и прогнозирования сердечного приступа	<a href="https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset?select=heart.csv">https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset?select=heart.csv</a>
3.	Диабет у индейцев пима	<a href="https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database">https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database</a>
4.	Данные по инсультам	<a href="https://www.kaggle.com/datasets/fedesorian/stroke-prediction-dataset">https://www.kaggle.com/datasets/fedesorian/stroke-prediction-dataset</a>
5.	Прогнозирование стоимости страховки	<a href="https://www.kaggle.com/datasets/mirichoi0218/insurance">https://www.kaggle.com/datasets/mirichoi0218/insurance</a>
6.	Продажи домов	<a href="https://www.kaggle.com/datasets/harlfoxem/housesalesprediction">https://www.kaggle.com/datasets/harlfoxem/housesalesprediction</a>
7.	Экономика стран	<a href="https://www.kaggle.com/datasets/pratik453609/economic-data-9-countries-19802020">https://www.kaggle.com/datasets/pratik453609/economic-data-9-countries-19802020</a>
8.	Заработная плата рабочих мест в области Data Science	<a href="https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries">https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries</a>
9.	Магазины	<a href="https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data">https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data</a>
10.	Оценки студентов на экзаменах	<a href="https://www.kaggle.com/datasets/spscientist/students-performance-in-exams">https://www.kaggle.com/datasets/spscientist/students-performance-in-exams</a>
11.	Цены на бриллианты	<a href="https://www.kaggle.com/datasets/nancyalaswad90/diamonds-prices">https://www.kaggle.com/datasets/nancyalaswad90/diamonds-prices</a>

№ варианта	Название набора данных	URL для скачивания
12.	Цены на кофе	<a href="https://www.kaggle.com/datasets/psycon/daily-coffee-price">https://www.kaggle.com/datasets/psycon/daily-coffee-price</a>
13.	Цены на акции	<a href="https://www.kaggle.com/datasets/nancyalawad90/yamana-gold-inc-stock-price">https://www.kaggle.com/datasets/nancyalawad90/yamana-gold-inc-stock-price</a>
14.	Цены на золото	<a href="https://www.kaggle.com/datasets/psycon/daily-gold-price-historical-data">https://www.kaggle.com/datasets/psycon/daily-gold-price-historical-data</a>
15.	Данные о населении	<a href="https://www.kaggle.com/datasets/muhammedtausif/world-population-by-countries">https://www.kaggle.com/datasets/muhammedtausif/world-population-by-countries</a>
16.	ЖЗЛ	<a href="https://www.kaggle.com/datasets/imoore/age-dataset">https://www.kaggle.com/datasets/imoore/age-dataset</a>
17.	Цены на автомобили	<a href="https://www.kaggle.com/datasets/deepcontractor/car-price-prediction-challenge">https://www.kaggle.com/datasets/deepcontractor/car-price-prediction-challenge</a>
18.	Цены на мобильные устройства	<a href="https://www.kaggle.com/datasets/midhunda/sl/mobile-price-hike-data">https://www.kaggle.com/datasets/midhunda/sl/mobile-price-hike-data</a>
19.	Данные о миллионерах	<a href="https://www.kaggle.com/datasets/surajjha101/forbes-billionaires-data-preprocessed">https://www.kaggle.com/datasets/surajjha101/forbes-billionaires-data-preprocessed</a>
20.	Он-лайн обучение	<a href="https://www.kaggle.com/datasets/shariful07/student-flexibility-in-online-learning">https://www.kaggle.com/datasets/shariful07/student-flexibility-in-online-learning</a>
21.	Данные о клиентах	<a href="https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis">https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis</a>
22.	Tesla Insider Trading	<a href="https://www.kaggle.com/datasets/ilyaryabov/tesla-insider-trading">https://www.kaggle.com/datasets/ilyaryabov/tesla-insider-trading</a>
23.	Продукты ЛЮ	<a href="https://www.kaggle.com/datasets/mohit2512/jio-mart-product-items">https://www.kaggle.com/datasets/mohit2512/jio-mart-product-items</a>
24.	НЛЮ	<a href="https://www.kaggle.com/datasets/sadeghjalilian/ufo-sightings-in-usa">https://www.kaggle.com/datasets/sadeghjalilian/ufo-sightings-in-usa</a>

№ варианта	Название набора данных	URL для скачивания
25.	Самые дорогие компании	<a href="https://www.kaggle.com/datasets/ankanhore/545/100-highest-valued-unicorns">https://www.kaggle.com/datasets/ankanhore/545/100-highest-valued-unicorns</a>

### Лабораторные работы по курсу «Методы искусственного интеллекта»

Для разработки веб-сервисов рекомендуется использовать Flask, но допускаются и другие варианты. Общая архитектура лабораторных работ представлена на рисунке 1.

Рассмотрим для примера разработку простого веб-сервиса. Пусть стоит задача разработать веб-сервис преобразования чисел в их словесное описание. Интерфейс показан на рисунке 2. Код реализации приведен в листинге 1.



Рисунок 1. Архитектура лабораторных работ

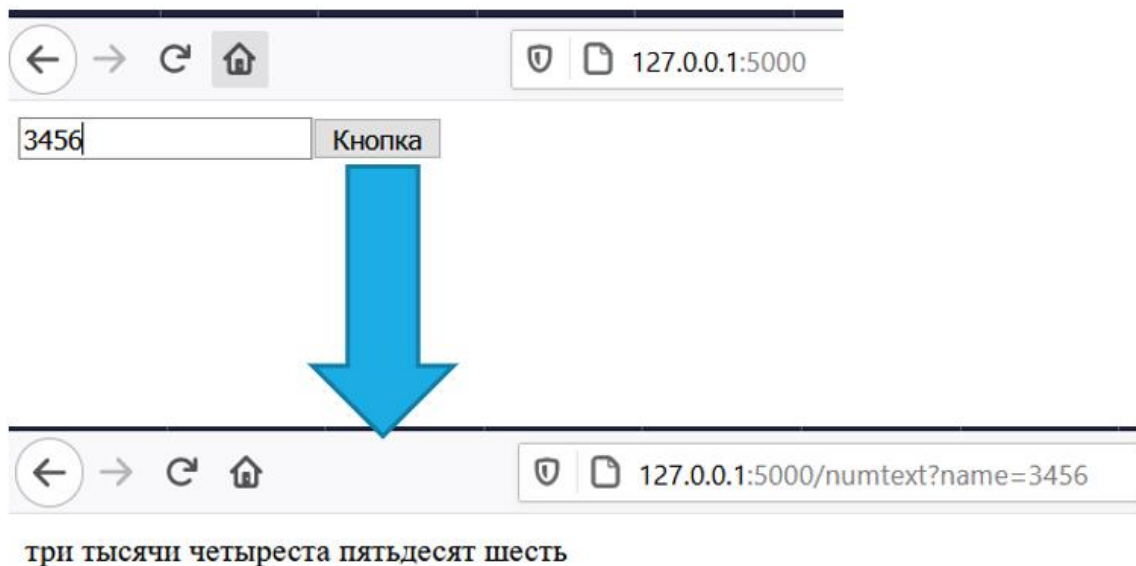


Рисунок 2. Интерфейс программы

Листинг 1. Код реализации веб-сервиса

```

from flask import Flask, redirect, url_for, request
from num2words import num2words

app=Flask(__name__)
@app.route("/")
def home():
    return "<html><form Action='http://127.0.0.1:5000/numtext'
Method=get>" \
        "<input type=text size=20 name=name>" \
        "<input type=submit value='Кнопка'>" \
        "</form></html>"

@app.route("/numtext", methods=['GET', 'POST'])
def numtext():
    data = request.args
    return num2words(data['name'], lang='ru')

if __name__=="__main__":
    app.run(debug=False)

```

Библиотеки flask и num2word у вас могут быть не установлены. Если вы работаете в среде PyCharm, то при наведении курсора появится «красная лампочка» и в ее меню будет команда импортирования, однако, она может выдать ошибку (рисунок 3). Чтобы провести импорт вам понадобится открыть командную строку, перейти в каталог с файлом pip.exe (на рисунке 3 – выделено) и повторить команду импорта (рисунок 4).



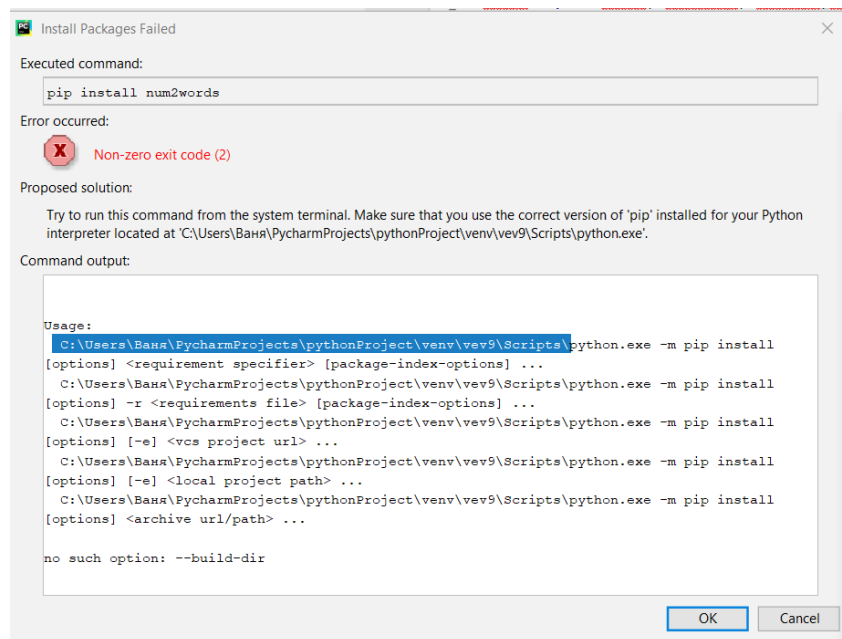


Рисунок 3. Ошибка импорта

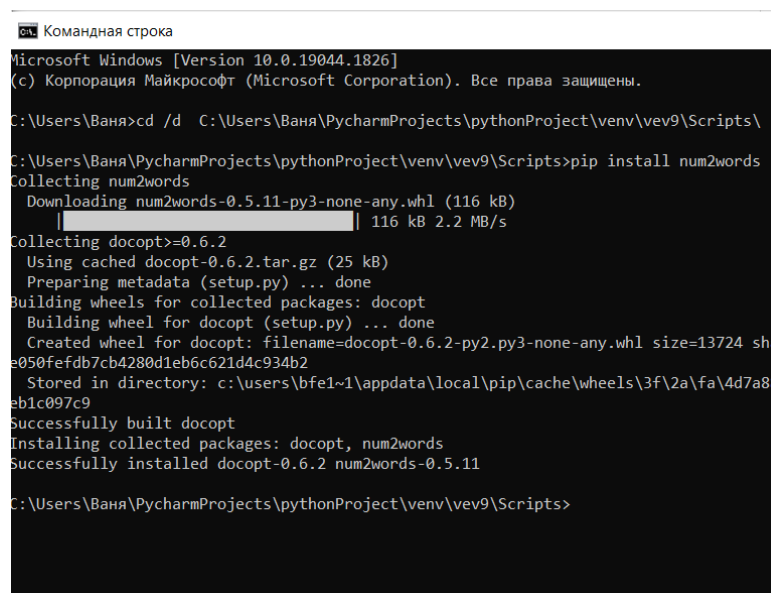


Рисунок 4. Импорт через командную строку

Остальной теоретический материал для выполнения лабораторных работ можно найти в лекционном курсе или здесь: [1, 2].

### Лабораторная работа 1. Веб-сервис для загрузки и представления данных

Для массива данных по варианту разработать простой веб-сервис на языке python по загрузке и выводу данных. Пользователь должен иметь

возможность указать, какой диапазон данных отразить (например, с 3 по 8ю строку, с 4й по 6й столбцы и т.д.). Вывести смысловое описание столбцов, указание типов данных столбцов, а также указание количества пустых и заполненных ячеек в них. Вывести описание самого набора данных, составленное на основе информации из таблицы, а также представленной разработчиком на странице для скачивания.

Контрольные вопросы по лабораторной работе:

1. Как разработать веб-сервис на python?
2. Каким образом вывести определенный диапазон данных?
3. Как считать данные из формата csv?

**Лабораторная работа 2. Веб-сервис для простейшего анализа данных**

Для массива данных по варианту добавить в веб-сервис функции решения указанных ниже задач (таблица 2). Дать смысловую интерпретацию полученным результатам. Например, пусть есть массив данных о ежедневных продажах акций в течение 10 лет. Стоит задача найти минимальную, максимальную и среднюю цену по годам. Вывод – таблица данных «год-цена мин., цена средняя, цена макс.». По ней возможный текст: «По проанализированным данным можно сделать вывод, что минимальная цена наблюдалась в таком-то году, максимальная – в таком-то году. Минимальная средняя цена была в таком-то году, максимальная – в таком-то».

Таблица 2. Задачи

№ варианта	Задания
1.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная и средняя скорости опасных объектов.</li> <li>• Задание 2. Минимальный, максимальный и средний размер опасных объектов.</li> </ul>

№ варианта	Задания
	<ul style="list-style-type: none"> <li>• Задание 3. Минимальная, максимальная и средняя скорости безопасных объектов.</li> <li>• Задание 4. Минимальный, максимальный и средний размер безопасных объектов.</li> </ul>
2.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальное, среднее, максимальное АД у мужчин</li> <li>• Задание 2. Минимальное, среднее, максимальное АД у женщин</li> <li>• Задание 3. Минимальное, среднее, максимальное значение сахара у мужчин</li> <li>• Задание 4. Минимальное, среднее, максимальное значение сахара у женщин</li> </ul>
3.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальное, среднее, максимальное АД по возрастным группам</li> <li>• Задание 2. Минимальное, среднее, максимальное значение инсулина по возрастным группам</li> <li>• Задание 3. Минимальное, среднее, максимальное значение ИМТ по возрастным группам</li> <li>• Задание 4. Минимальное, среднее, максимальное значение количества беременностей по возрастным группам</li> </ul>
4.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальное, среднее, максимальное значение возраста у холостых</li> <li>• Задание 2. Минимальное, среднее, максимальное значение возраста у женатых</li> <li>• Задание 3. Минимальное, среднее, максимальное значение ИМТ у мужчин</li> </ul>

№ варианта	Задания
	<ul style="list-style-type: none"> <li>• Задание 4. Минимальное, среднее, максимальное значение ИМТ у женщин</li> </ul>
5.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальное, среднее, максимальное значение возраста у курильщиков</li> <li>• Задание 2. Минимальное, среднее, максимальное значение возраста у некурящих</li> <li>• Задание 3. Минимальное, среднее, максимальное значение ИМТ у мужчин</li> <li>• Задание 4. Минимальное, среднее, максимальное значение ИМТ у женщин</li> </ul>
6.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена для домов с больше, чем двумя ваннами.</li> <li>• Задание 2. Минимальная, максимальная, средняя цена у домов, сгруппированных по количеству комнат.</li> <li>• Задание 3. Минимальная, максимальная, средняя цена у домов, сгруппированных по количеству этажей.</li> <li>• Задание 4. Минимальная, максимальная, средняя цена у домов на набережных</li> </ul>
7.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена на нефть по странам.</li> <li>• Задание 2. Минимальная, максимальная, средняя цена на нефть по годам.</li> <li>• Задание 3. Минимальный, максимальный, средний уровень безработицы по странам.</li> <li>• Задание 4. Минимальный, максимальный, средний уровень безработицы по годам.</li> </ul>

№ варианта	Задания
8.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя зарплата, сгруппированная по уровням опыта</li> <li>• Задание 2. Минимальная, максимальная, средняя зарплата, сгруппированная по должности</li> <li>• Задание 3. Минимальная, максимальная, средняя зарплата, сгруппированная по размеру компании</li> <li>• Задание 4. Минимальная, максимальная, средняя зарплата, сгруппированная по году работы</li> </ul>
9.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальный, максимальный, средний объем доступных товаров, сгруппированный по площади магазина.</li> <li>• Задание 2. Минимальный, максимальный, средний пассажиропоток, сгруппированный по площади магазина.</li> <li>• Задание 3. Минимальный, максимальный, средний объем выручки, сгруппированный по площади магазина.</li> <li>• Задание 4. Минимальный, максимальный, средний объем выручки, сгруппированный по пассажиропотоку.</li> </ul>
10.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальный, максимальный, средний балл по математике в разрезе гендерной принадлежности.</li> <li>• Задание 2. Минимальный, максимальный, средний балл по математике в разрезе образования.</li> <li>• Задание 3. Минимальный, максимальный, средний балл по математике в разрезе этнической принадлежности.</li> <li>• Задание 4. Минимальный, максимальный, средний балл по математике в разрезе подготовительных курсов.</li> </ul>
11.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена, сгруппированная по огранке.</li> </ul>

№ варианта	Задания
	<ul style="list-style-type: none"> <li>• Задание 2. Минимальная, максимальная, средняя цена, сгруппированная по каратам.</li> <li>• Задание 3. Минимальная, максимальная, средняя цена, сгруппированная по прозрачности.</li> <li>• Задание 4. Минимальная, максимальная, средняя цена, сгруппированная по цвету.</li> </ul>
12.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена открытия летом (во все года)</li> <li>• Задание 2. Минимальная, максимальная, средняя цена открытия зимой по годам</li> <li>• Задание 3. Минимальная, максимальная, средняя цена открытия за выбранный год</li> <li>• Задание 4. Минимальная, максимальная, средняя цена открытия по годам</li> </ul>
13.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена открытия летом (во все года)</li> <li>• Задание 2. Минимальная, максимальная, средняя цена открытия зимой по годам</li> <li>• Задание 3. Минимальная, максимальная, средняя цена открытия за выбранный год</li> <li>• Задание 4. Минимальная, максимальная, средняя цена открытия по годам</li> </ul>
14.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена открытия летом (во все года)</li> <li>• Задание 2. Минимальная, максимальная, средняя цена открытия зимой по годам</li> </ul>

№ варианта	Задания
	<ul style="list-style-type: none"> <li>• Задание 3. Минимальная, максимальная, средняя цена открытия за выбранный год</li> <li>• Задание 4. Минимальная, максимальная, средняя цена открытия по годам</li> </ul>
15.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя численность по континентам.</li> <li>• Задание 2. Минимальная, максимальная, средняя численность на островах.</li> <li>• Задание 3. Минимальная, максимальная, средняя численность, сгруппированная по площади страны</li> </ul>
16.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя длительность жизни в разрезе гендера</li> <li>• Задание 2. Минимальная, максимальная, средняя длительность жизни в разрезе вида деятельности</li> <li>• Задание 3. Минимальный, максимальный, средний год рождения в разрезе вида деятельности</li> <li>• Задание 4. Минимальное, максимальное, среднее количество деятелей в разрезе веков</li> </ul>
17.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена в разрезе марок</li> <li>• Задание 2. Минимальная, максимальная, средняя цена в разрезе типа кузова</li> <li>• Задание 3. Минимальная, максимальная, средняя цена в разрезе годов производства</li> <li>• Задание 4. Минимальная, максимальная, средняя цена в разрезе типа топлива</li> </ul>

№ варианта	Задания
18.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена в разрезе производителя</li> <li>• Задание 2. Минимальная, максимальная, средняя цена в разрезе цвета</li> <li>• Задание 3. Минимальная, максимальная, средняя цена в разрезе размеров монитора</li> <li>• Задание 4. Минимальная, максимальная, средняя цена в разрезе типа процессора</li> </ul>
19.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальный, максимальный, средний размер собственного капитала в разрезе страны</li> <li>• Задание 2. Минимальный, максимальный, средний размер собственного капитала, сгруппированный по возрасту.</li> <li>• Задание 3. Минимальный, максимальный, средний размер собственного капитала в разрезе вида деятельности</li> <li>• Задание 4. Минимальный, максимальный, средний размер собственного капитала в разрезе источника богатства</li> </ul>
20.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальный, максимальный, средний возраст в разрезе гендера</li> <li>• Задание 2. Минимальный, максимальный, средний возраст в разрезе уровня гибкости</li> <li>• Задание 3. Минимальный, максимальный, средний возраст в разрезе уровня обучения</li> <li>• Задание 4. Минимальный, максимальный, средний возраст в разрезе устройства, с которого человек работает</li> </ul>
21.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя сумма потраченная на вино, в разрезе возраста (считаем от текущей даты).</li> </ul>



№ варианта	Задания
	<ul style="list-style-type: none"> <li>• Задание 2. Минимальная, максимальная, средняя сумма потраченная на вино, в разрезе образования.</li> <li>• Задание 3. Минимальная, максимальная, средняя сумма потраченная на вино, в разрезе социального статуса (женат, разведен и т.д.).</li> <li>• Задание 4. Минимальная, максимальная, средняя сумма потраченная на вино, в разрезе сгруппированного дохода семьи клиента.</li> </ul>
22.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя сумма сделки по персоналиям.</li> <li>• Задание 2. Минимальная, максимальная, средняя сумма сделки по статусу персоналии в компании.</li> <li>• Задание 3. Минимальная, максимальная, средняя сумма сделки по типу транзакции.</li> <li>• Задание 4. Минимальная, максимальная, средняя сумма сделки по месяцам.</li> </ul>
23.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальная, максимальная, средняя цена в разрезе категорий</li> <li>• Задание 2. Минимальная, максимальная, средняя цена в разрезе подкатегорий</li> <li>• Задание 3. Минимальная, максимальная, средняя цена за килограмм</li> </ul>
24.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальное, максимальное, среднее количество по городам.</li> <li>• Задание 2. Минимальное, максимальное, среднее количество по штатам.</li> </ul>

№ варианта	Задания
	<ul style="list-style-type: none"> <li>• Задание 3. Минимальная, максимальная, средняя длительность.</li> <li>• Задание 4. Минимальная, максимальная, средняя разница во времени между контактом и публикацией данных о нем.</li> </ul>
25.	<ul style="list-style-type: none"> <li>• Задание 1. Минимальное, максимальное, среднее общее финансирование по странам.</li> <li>• Задание 2. Минимальное, максимальное, среднее общее финансирование по году основания.</li> <li>• Задание 3. Минимальное, максимальное, среднее общее финансирование по количеству служащих.</li> <li>• Задание 4. Минимальное, максимальное, среднее общее финансирование по городам.</li> </ul>

Контрольные вопросы по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. Как выполнить группировку данных?
3. Что такое семантическая интерпретация результатов?

**Лабораторная работа 3. Веб-сервис для визуализации данных**

Дополнить свой массив данных на 10% усредненными значениями для числовых столбцов и наиболее часто встречающимися для текстовых. Пересчитать задания предыдущей лабораторной работы и визуализировать прошлые и новые результаты с помощью графиков. Ответить на вопрос – привело ли увеличение объема данных к их искажению?

Контрольные вопросы по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. Как вы дополнили массив данных?
3. С помощью чего вы строили графики?

#### **Лабораторная работа 4. Веб-сервис «Фильтр Блума» и интеграция сервисов**

Для разработанного в первой лабораторной работе описания данных сделать фильтр Блума (сдается каждым студентом единолично) и интегрировать разработанные в рамках группы сервисы в единую систему, работающую следующим образом. Пользователь задает ключевые слова для поиска массива данных. Например, «дома». Сначала поиск происходит внутри фильтра, если есть смысл искать, то поиск происходит по некой структуре данных, которую вам необходимо разработать, и пользователя перекидывает на страницу нужного веб-сервиса, если он единственный по описанию или на страницу выбора, если под описание подходит несколько сервисов. В дальнейшем все лабораторные сдаются через эту систему, работы по созданию которой студенты сами распределяют внутри группы. Сданные ранее лабораторные также должны быть интегрированы в эту систему.

##### Контрольные вопросы по лабораторной работе:

1. В чем особенность и какова сфера применения фильтра Блума?
2. Какую хеш-функцию вы выбрали и почему?
3. Как вы интегрировали свой сервис в общую систему?

#### **Лабораторная работа 5. Веб-сервис «Парная линейная регрессия»**

Запрограммировать парную линейную регрессию для задачи по варианту (таблица 3) на 99% ваших данных. Оценить, насколько хорошо она подходит для решения данной задачи (для каких-то задач она может выдавать совершенно некачественные результаты) на оставшемся проценте. При необходимости использовать стандартизацию и нормализацию данных.

Таблица 3. Задачи

№ варианта	Задания
1.	Зависимость скорости от размера.
2.	Зависимость АД от уровня сахара.
3.	Зависимость АД от ИМТ.
4.	Зависимость ИМТ от возраста.
5.	Зависимость ИМТ от возраста.
6.	Зависимость цены от этажности
7.	Зависимость уровня безработицы от цены на нефть
8.	Зависимость зарплаты от размера компании
9.	Зависимость выручки от площади магазина
10.	Зависимость балла по математике от образования.
11.	Зависимость цены от каратов
12.	Зависимость цены закрытия от объема
13.	Зависимость цены закрытия от цены открытия
14.	Зависимость цены закрытия от цены открытия
15.	Зависимость численности от площади
16.	Зависимость длительности жизни от века
17.	Зависимость цены от года производства
18.	Зависимость цены от размера монитора
19.	Зависимость собственного капитала от возраста
20.	Зависимость уровня гибкости от возраста
21.	Зависимость трат на вино от дохода семьи клиента.
22.	Зависимость типа транзакции от суммы сделки
23.	Зависимость цены от категории
24.	Зависимость длительности от времени года
25.	Зависимость финансирования от количества служащих

### Контрольные вопросы по лабораторной работе:

1. Какие еще задачи можно решить с помощью регрессии на вашем наборе данных?
2. Как вы запрограммировали парную линейную регрессию?
3. Что вы можете сказать о качестве решения?

### **Лабораторная работа 6. Веб-сервис «Дерево решений»**

Запрограммировать дерево решений как минимум на 25 строках ваших данных для задачи по варианту (таблица 4). Оценить на 5 записях из оставшихся данных, насколько хорошо оно подходит для решения данной задачи.

Таблица 4. Задачи

№ варианта	Задания
1.	Зависимость опасности объекта от размера и скорости
2.	Зависимость АД от пола и возраста
3.	Зависимость АД от возраста и ИМТ
4.	Зависимость ИМТ от пола и возраста
5.	Значение ИМТ от пола и курения
6.	Зависимость цены от этажности и количества комнат
7.	Зависимость цены на нефть от страны и года
8.	Зависимость зарплаты от уровня опыта и должности
9.	Зависимость выручки от пассажиропотока и площади магазина
10.	Зависимость бала по математике от гендерной и этнической принадлежности
11.	Зависимость цены от огранки и прозрачности
12.	Зависимость цены открытия от времени года
13.	Зависимость цены открытия от времени года

№ варианта	Задания
14.	Зависимость цены открытия от времени года
15.	Зависимость численности от континента
16.	Зависимость длительности жизни от пола и вида деятельности
17.	Зависимость цены от марки и года производства
18.	Зависимость цены от производителя и монитора
19.	Зависимость собственного капитала от страны и источника финансирования
20.	Зависимость уровня гибкости от возраста и устройства, с которого человек работает
21.	Траты на вино в зависимости от возраста и образования
22.	Зависимость суммы сделки от типа транзакции и месяца.
23.	Зависимость цены от категории и подкатегории
24.	Зависимость длительности контакта от времени года и города
25.	Зависимость финансирования от страны и количества служащих

Контрольные вопросы по лабораторной работе:

1. Какие еще задачи можно решить с помощью дерева решений на вашем наборе данных?
2. Как вы запрограммировали дерево решений?
3. Что вы можете сказать о качестве решения?

**Лабораторная работа 7. Веб-сервис «Кластеризация»**

Запрограммировать алгоритм кластеризации для задачи по варианту (таблица 5). Оценить, насколько хорошо он подходит для решения данной задачи. Алгоритм студент выбирает сам, библиотечными реализациями не пользуясь.

Таблица 5. Задачи

№ варианта	Параметры кластеризации
1.	Скорость и размер.
2.	АД и уровень сахара.
3.	АД и ИМТ.
4.	ИМТ и возраст.
5.	ИМТ и возраст.
6.	Цена и этажность
7.	Уровень безработицы и цены на нефть
8.	Зарплата и размер компании
9.	Выручка и площадь магазина
10.	Баллы по математике и образование.
11.	Цена и караты
12.	Цена закрытия и объем
13.	Цены закрытия и открытия
14.	Цены закрытия и открытия
15.	численности и площади
16.	длительности жизни и века
17.	цены и года производства
18.	цены и размера монитора
19.	собственного капитала и возраста
20.	уровня гибкости и возраста
21.	трат на вино и дохода семьи клиента.
22.	типа транзакции и суммы сделки
23.	цены и категории
24.	длительности и времени года
25.	финансирования и количества служащих

### Контрольные вопросы по лабораторной работе:

1. Чем кластеризация отличается от классификации?
2. Почему для кластеризации важно правильно выбрать критерий качества и меру расстояний?
3. Что вы можете сказать о качестве решения?

### **Лабораторная работа 8. Генетический алгоритм**

1. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый дешевый рацион из  $k$  наименований, удовлетворяющий заданным медицинским нормам для каждой из  $m$  характеристик.

2. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный

маршрут, так, чтобы все города получили нужный им объем продуктов без сильного его превышения, а транспортные расходы были минимальными.

3. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый лучший по характеристикам рацион из  $k$  наименований, удовлетворяющий заданным ценовым рамкам. Лучшим считается рацион с минимальным отклонением от нормы.

4. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются.



Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов с минимальным его превышением, а транспортные расходы укладывались в определенные рамки.

5. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  полей для  $k$  культур для посева. Для каждого поля известна характеристика урожайности каждой из  $k$  культур, а для каждой культуры – его закупочная стоимость. Необходимо получить самый лучший урожай за наименьшую стоимость.

6. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый дешевый рацион из  $k$  наименований, удовлетворяющий заданным медицинским нормам для каждой из  $m$  характеристик.

7. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный

маршрут, так, чтобы все города получили нужный им объем продуктов без сильного его превышения, а транспортные расходы были минимальными.

8. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый лучший по характеристикам рацион из  $k$  наименований, удовлетворяющий заданным ценовым рамкам. Лучшим считается рацион с минимальным отклонением от нормы.

9. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов с минимальным его превышением, а транспортные расходы укладывались в определенные рамки.

10. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  полей для  $i$  и  $k$  культур для посева. Для каждого поля известна характеристика урожайности каждой из  $k$  культур, а для каждой культуры – его закупочная стоимость. Необходимо получить самый лучший урожай за наименьшую стоимость.

11. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый дешевый рацион из  $k$  наименований, удовлетворяющий заданным медицинским нормам для каждой из  $m$  характеристик.

12. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный

маршрут, так, чтобы все города получили нужный им объем продуктов без сильного его превышения, а транспортные расходы были минимальными.

13. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый лучший по характеристикам

рацион из  $k$  наименований, удовлетворяющий заданным ценовым рамкам. Лучшим считается рацион с минимальным отклонением от нормы.

14. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов с минимальным его превышением, а транспортные расходы укладывались в определенные рамки.

15. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  полей для  $i$  и  $k$  культур для посева. Для каждого поля известна характеристика урожайности каждой из  $k$  культур, а для каждой культуры – его закупочная стоимость. Необходимо получить самый лучший урожай за наименьшую стоимость.

16. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый дешевый рацион из  $k$  наименований, удовлетворяющий заданным медицинским нормам для каждой из  $m$  характеристик.

17. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов без сильного его превышения, а транспортные расходы были минимальными.

18. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано

$N$  наименований продуктов, для каждого из которых известно  $m$  характеристик. Необходимо получить самый лучший по характеристикам рацион из  $k$  наименований, удовлетворяющий заданным ценовым рамкам. Лучшим считается рацион с минимальным отклонением от нормы.

19. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $n$  пунктов производства продуктов и  $k$  городов, которые в них нуждаются. Каждый город может потребить  $x$  продуктов, а каждый пункт произвести  $y$  продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов с минимальным его превышением, а транспортные расходы укладывались в определенные рамки.

20. На языке Python разработайте скрипт, который с помощью генетического алгоритма и полного перебора решает следующую задачу. Дано  $N$  полей для  $i$  и  $k$  культур для посева. Для каждого поля известна характеристика урожайности каждой из  $k$  культур, а для каждой культуры – его закупочная стоимость. Необходимо получить самый лучший урожай за наименьшую стоимость.

#### Контрольные вопросы по лабораторной работе:

1. Какова структура хромосомы для вашей задачи?
2. Какую функцию приспособленности вы выбрали и почему?
3. Что такое кроссовер и мутация?

### **Лабораторная работа 9. Нечеткая логика**

1. На языке Python разработайте скрипт, позволяющий задать нечеткое множество с треугольной функцией принадлежности и отобразить его название, параметры, а также степень принадлежности вводимого пользователем объекта.

2. На языке Python разработайте скрипт, позволяющий задать нечеткое множество с трапециевидной функцией принадлежности и отобразить его

параметры, а также степень принадлежности вводимого пользователем объекта.

3. На языке Python разработайте скрипт, позволяющий выполнить операцию пересечения заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – пересечение данных нечетких множеств.

4. На языке Python разработайте скрипт, позволяющий выполнить операцию пересечения заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – пересечение данных нечетких множеств.

5. На языке Python разработайте скрипт, позволяющий выполнить операцию объединения заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – объединение данных нечетких множеств.

6. На языке Python разработайте скрипт, позволяющий выполнить операцию объединения заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – объединение данных нечетких множеств.

7. На языке Python разработайте скрипт, позволяющий выполнить операцию дополнения, заданного пользователем нечеткого множества с треугольной функцией принадлежности. Входными данными будут параметры функции принадлежности и четкие объекты множества. Выходными – дополнение нечеткого множества.

8. На языке Python разработайте скрипт, позволяющий выполнить операцию дополнения, заданного пользователем нечеткого множества с трапециевидной функцией принадлежности. Входными данными будут

параметры функции принадлежности и четкие объекты множества. Выходными – дополнение нечеткого множества.

9. На языке Python разработайте скрипт, позволяющий выполнить операцию импликации заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – результат импликации данных нечетких множеств. Причем, результат вывести через лингвистические переменные. Импликацию моделировать минимумом.

10. На языке Python разработайте скрипт, позволяющий выполнить операцию импликации заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – результат импликации данных нечетких множеств. Импликацию моделировать минимумом.

11. На языке Python разработайте скрипт, позволяющий задать нечеткое множество с треугольной функцией принадлежности и отобразить его название, параметры, а также степень принадлежности вводимого пользователем объекта.

12. На языке Python разработайте скрипт, позволяющий задать нечеткое множество с трапециевидной функцией принадлежности и отобразить его параметры, а также степень принадлежности вводимого пользователем объекта.

13. На языке Python разработайте скрипт, позволяющий выполнить операцию пересечения заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – пересечение данных нечетких множеств.

14. На языке Python разработайте скрипт, позволяющий выполнить операцию пересечения заданных пользователем нечетких множеств с

трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – пересечение данных нечетких множеств.

15. На языке Python разработайте скрипт, позволяющий выполнить операцию объединения заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – объединение данных нечетких множеств.

16. На языке Python разработайте скрипт, позволяющий выполнить операцию объединения заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – объединение данных нечетких множеств.

17. На языке Python разработайте скрипт, позволяющий выполнить операцию дополнения, заданного пользователем нечеткого множества с треугольной функцией принадлежности. Входными данными будут параметры функции принадлежности и четкие объекты множества. Выходными – дополнение нечеткого множества.

18. На языке Python разработайте скрипт, позволяющий выполнить операцию дополнения, заданного пользователем нечеткого множества с трапециевидной функцией принадлежности. Входными данными будут параметры функции принадлежности и четкие объекты множества. Выходными – дополнение нечеткого множества.

19. На языке Python разработайте скрипт, позволяющий выполнить операцию импликации заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – результат импликации данных нечетких множеств. Причем, результат вывести через лингвистические переменные. Импликацию моделировать минимумом.

20. На языке Python разработайте скрипт, позволяющий выполнить операцию импликации заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – результат импликации данных нечетких множеств. Импликацию моделировать минимумом.

Контрольные вопросы по лабораторной работе:

1. Дайте определение нечеткому множеству.
2. Какие способы задания функций принадлежности вы знаете?
3. Какую операцию вы реализовали в своей лабораторной работе?

**Лабораторная работа 10. Лингвистические переменные и шкалы**

Разработать программное обеспечение на языке python, позволяющее задавать с использованием треугольных или трапециевидных функций принадлежности пользовательскую лингвистическую шкалу для решения задачи по варианту (таблица 6). Количество оценок в шкале и параметры функций принадлежности должны задаваться по умолчанию, но редактироваться пользователем. Осуществить графическое отображение функций принадлежности меток шкалы, выделив каждую отдельным цветом.

Таблица 6. Варианты заданий

Вариант	Назначение шкалы
1.	Оценка эффективности продаж
2.	Оценка эффективности операций с валютой
3.	Оценка прибыли
4.	Оценка затрат на оплату труда
5.	Оценка рентабельности



Вариант	Назначение шкалы
6.	Оценка объемов продаж
7.	Оценка объемов закупок
8.	Оценка загрузки сервера и сети
9.	Оценка уровня зарплаты
10.	Оценка убытков
11.	Оценка эффективности продаж
12.	Оценка прибыли
13.	Оценка затрат на оплату труда
14.	Оценка объемов продаж
15.	Оценка объемов закупок
16.	Оценка загрузки сервера и сети
17.	Оценка уровня зарплаты
18.	Оценка убытков
19.	Оценка прибыли
20.	Оценка затрат на оплату труда
21.	Оценка объемов продаж
22.	Оценка объемов закупок
23.	Оценка загрузки сервера и сети
24.	Оценка уровня зарплаты
25.	Оценка убытков

### Контрольные вопросы по лабораторной работе:

1. Дайте определение лингвистической переменной.
2. Дайте определение лингвистической шкале.
3. Какие функции принадлежности вы использовали?

### **Лабораторная работа 11. Нечеткая кластеризация объектов**

1. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере заработной платы  $n$  людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

2. На языке Python разработайте скрипт, кластеризующий загруженные данные о возрасте  $n$  людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

3. На языке Python разработайте скрипт, кластеризующий загруженные данные о стоимости  $n$  автомобилей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

4. На языке Python разработайте скрипт, кластеризующий загруженные данные о росте  $n$  людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

5. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере затрат на производство  $n$  продуктов на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

6. На языке Python разработайте скрипт, кластеризующий загруженные данные о длительности перелетов до  $n$  пунктов на определенные им кластеры,

обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

7. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере затрат на связь среди  $n$  отделов на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

8. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере площадей  $n$  квартир на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

9. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере урожая  $n$  сельхоз культур на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

10. На языке Python разработайте скрипт, кластеризующий загруженные данные о весе  $n$  людей на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

11. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере заработной платы  $n$  людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

12. На языке Python разработайте скрипт, кластеризующий загруженные данные о возрасте  $n$  людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

13. На языке Python разработайте скрипт, кластеризующий загруженные данные о стоимости  $n$  автомобилей на определенные им кластеры,

обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

14. На языке Python разработайте скрипт, кластеризующий загруженные данные о росте  $n$  людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

15. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере затрат на производство  $n$  продуктов на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

16. На языке Python разработайте скрипт, кластеризующий загруженные данные о длительности перелетов до  $n$  пунктов на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

17. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере затрат на связь среди  $n$  отделов на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

18. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере площадей  $n$  квартир на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

19. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере урожая  $n$  сельхоз культур на определенные им кластеры, обозначенные определенными в

программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

20. На языке Python разработайте скрипт, кластеризующий загруженные данные о весе  $n$  людей на определенные им кластеры, обозначенные

определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

Контрольные вопросы по лабораторной работе:

1. Чем нечеткая кластеризация отличается от четкой?
2. Какие метки вы задали для демонстрационного примера?
3. Какие функции принадлежности вы использовали?

**Лабораторная работа 12. Нечеткий логический вывод**

Реализовать систему нечеткого логического вывода по варианту (таблица 7). В базе правил должно быть не меньше десяти правил.

Таблица 7. Варианты заданий

Остаток от деления варианта на 5	Входные переменные	Выходная переменная
0.	Стаж, Количество положительных отзывов	Уровень зарплаты
1.	Часы эксплуатации устройства, Срок последней проверки	Риск поломки
2.	Расстояние до Земли, Размер небесного тела, Скорость движения к Земле	Степень опасности
3.	Индекс массы тела, Уровень глюкозы	Вероятность диабета
4.	Скорость, расстояние для преодоления	Объем топлива

Контрольные вопросы по лабораторной работе:

1. Что такое нечеткий логический вывод?

2. Какую схему вы использовали?
3. Какие функции принадлежности вы использовали?

### Лабораторная работа 13. Анализ текстов

Выполните анализ текста из сборника трудов ИМАП <http://venec.ulstu.ru/lib/go.php?id=8676> по варианту (таблица 8, порядковый номер статьи в оглавлении – номер варианта). Удалите стоп-слова. Для статистического анализа не забудьте выполнить морфологическую предобработку текста.

Таблица 8. Варианты заданий

Остаток от деления варианта на 5	Статистический анализ	Морфологический анализ
0	Постройте частотный словарь.	Программно определите количество глаголов.
1	Выявите наиболее статистически значимые двусловия по методу Mutual Information.	Программно определите количество существительных женского рода.
2	Выявите наиболее статистически значимые двусловия по методу T-Score.	Программно определите количество прилагательных.
3	Выявите наиболее статистически	Программно определите количество существительных,

Остаток от деления варианта на 5	Статистический анализ	Морфологический анализ
	значимые двусловия по методу Log-Likelihood.	стоящих во множественном числе.
4	Выявите наиболее статистически значимые двусловия по методу TF×IDF	Программно определите количество глаголов, стоящих в прошедшем времени.

Контрольные вопросы по лабораторной работе:

1. Что такое статистический анализ текстов?
2. Что такое морфологический анализ текстов?
3. Что такое стоп-слова?

**Лабораторная работа 14. Нейронные сети**

Ознакомившись с пунктом «Реализация алгоритма обучения нейронной сети» ([1], страницы 223-228), разработайте нейронную сеть и реализуйте алгоритм ее обучения. Количество слоев – три плюс остаток от деления номера варианта на 3. В первом слое – три нейрона. В каждом промежуточном – на один нейрон больше, чем в предыдущем. Выходной слой имеет один нейрон. Тестовые и тренировочные данные взять из источника, как и функцию активации и количество эпох обучения.

Контрольные вопросы по лабораторной работе:

1. Что такое градиент?
2. Что такое обратное распространение ошибки?
3. Что такое эпоха обучения?

## Лабораторные работы по курсу «Машинное обучение»

В лабораторных 4-6 студентам необходимо показать навык самостоятельного формулирования задач для заданных данных (предполагается, что он получен ими в курсе «Методы искусственного интеллекта»). Необходимый теоретический материал для лабораторных работ можно найти в лекционном курсе или здесь: [1].

### Лабораторная работа 1. Работа с типовыми наборами данных и различными моделями

Используя код из пункта «Регуляризация и сеть прямого распространения» из [1] (стр. 228), сгенерируйте определенный тип данных и сравните на нем 3 модели (по варианту). Постройте графики, отобразите качество моделей, объясните полученные результаты.

1. Данные: `make_moons (noise=0.3, random_state=rs)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 3)
- Гребневую полиномиальную регрессию (со степенью 3,  $\alpha = 1.0$ )

2. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 3)
- Гребневую полиномиальную регрессию (со степенью 3,  $\alpha = 1.0$ )

3. Данные: `make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 3)
- Гребневую полиномиальную регрессию (со степенью 3,  $\alpha = 1.0$ )



4. Данные: `make_moons (noise=0.3, random_state=rs)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 4)
- Гребневую полиномиальную регрессию (со степенью 4,  $\alpha = 1.0$ )

5. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 4)
- Гребневую полиномиальную регрессию (со степенью 4,  $\alpha = 1.0$ )

6. Данные: `make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 4)
- Гребневую полиномиальную регрессию (со степенью 4,  $\alpha = 1.0$ )

7. Данные: `make_moons (noise=0.3, random_state=rs)`

Модели:

- Персептрон
- Многослойный персептрон с 10-ю нейронами в скрытом слое ( $\alpha = 0.01$ )
- Многослойный персептрон со 100-а нейронами в скрытом слое ( $\alpha = 0.01$ )

8. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`

Модели:

- Персептрон
- Многослойный персептрон с 10-ю нейронами в скрытом слое ( $\alpha = 0.01$ )
- Многослойный персептрон со 100-а нейронами в скрытом слое ( $\alpha = 0.01$ )

9. Данные: `make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`

Модели:

- Перцептрон
- Многослойный перцептрон с 10-ю нейронами в скрытом слое ( $\alpha = 0.01$ )
- Многослойный перцептрон со 100-а нейронами в скрытом слое ( $\alpha = 0.01$ )

10. Данные: `make_moons (noise=0.3, random_state=rs)`

Модели:

- Линейную регрессию
- Многослойный перцептрон с 10-ю нейронами в скрытом слое ( $\alpha = 0.01$ )
- Многослойный перцептрон со 100-а нейронами в скрытом слое ( $\alpha = 0.01$ )

11. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 4)
- Перцептрон

12. Данные: `make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`

Модели:

- Линейную регрессию
- Перцептрон
- Гребневую полиномиальную регрессию (со степенью 4,  $\alpha = 1.0$ )

13. Данные: `make_moons (noise=0.3, random_state=rs)`

Модели:

- Линейную регрессию
- Полиномиальную регрессию (со степенью 3)

- Многослойный перцептрон со 100-а нейронами в скрытом слое ( $\alpha = 0.01$ )
14. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`
- Модели:
- Линейную регрессию
  - Перцептрон
  - Гребневую полиномиальную регрессию (со степенью 4,  $\alpha = 1.0$ )
15. Данные: `make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`
- Модели:
- Линейную регрессию
  - Полиномиальную регрессию (со степенью 4)
  - Перцептрон
16. Данные: `make_moons (noise=0.3, random_state=rs)`
- Модели:
- Линейную регрессию
  - Многослойный перцептрон с 10-ю нейронами в скрытом слое ( $\alpha = 0.01$ )
  - Перцептрон
17. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`
- Модели:
- Линейную регрессию
  - Полиномиальную регрессию (со степенью 3)
  - Многослойный перцептрон со 100-а нейронами в скрытом слое ( $\alpha = 0.01$ )
18. Данные: `make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`
- Модели:
- Линейную регрессию
  - Перцептрон

· Гребневую полиномиальную регрессию (со степенью 3,  $\alpha = 1.0$ )

19. Данные: `make_moons (noise=0.3, random_state=rs)`

Модели:

· Линейную регрессию

· Полиномиальную регрессию (со степенью 5)

· Гребневую полиномиальную регрессию (со степенью 5,  $\alpha = 1.0$ )

20. Данные: `make_circles (noise=0.2, factor=0.5, random_state=rs)`

Модели:

· Линейную регрессию

· Полиномиальную регрессию (со степенью 5)

· Гребневую полиномиальную регрессию (со степенью 5,  $\alpha = 1.0$ )

21. Данные: `make_classification (n_samples=500, n_features=2,`

`n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)`

Модели:

· Линейную регрессию

· Полиномиальную регрессию (со степенью 5)

· Гребневую полиномиальную регрессию (со степенью 5,  $\alpha = 1.0$ )

Контрольные вопросы по лабораторной работе:

1. Зачем нужны типовые наборы данных?

2. По каким критериям вы оценивали работы моделей?

3. Какая модель показала лучшие результаты на вашем наборе данных?

## **Лабораторная работа 2. Ранжирование признаков**

Используя код из [1] (пункт «Решение задачи ранжирования признаков», стр. 205), выполните ранжирование признаков с помощью указанных по варианту моделей. Отобразите получившиеся значения\оценки каждого признака каждым методом\моделью и среднюю оценку. Проведите анализ получившихся результатов. Какие четыре признака оказались самыми

важными по среднему значению? (Названия\индексы признаков и будут ответом на задание).

1. Линейная регрессия (LinearRegression), Случайное Лассо (RandomizedLasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE)
2. Линейная регрессия (LinearRegression), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Сокращение признаков Случайными деревьями (Random Forest Regressor)
3. Линейная регрессия (LinearRegression) , Сокращение признаков Случайными деревьями (Random Forest Regressor), Линейная корреляция (f\_regression)
4. Гребневая регрессия (Ridge), Случайное Лассо (RandomizedLasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE)
5. Гребневая регрессия (Ridge), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Сокращение признаков Случайными деревьями (Random Forest
6. Гребневая регрессия (Ridge), Сокращение признаков Случайными деревьями (Random Forest Regressor), Линейная корреляция (f\_regression)
7. Лассо (Lasso), Случайное Лассо (RandomizedLasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE)
8. Лассо (Lasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Сокращение признаков Случайными деревьями (Random Forest Regressor)
9. Лассо (Lasso),Сокращение признаков Случайными деревьями (Random Forest Regressor), Линейная корреляция (f\_regression)

10. Линейная регрессия (LinearRegression), Лассо (Lasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE)
11. Гребневая регрессия (Ridge), Случайное Лассо (RandomizedLasso), Сокращение признаков Случайными деревьями (Random Forest Regressor)
12. Лассо (Lasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Линейная корреляция (f\_regression)
13. Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Сокращение признаков Случайными деревьями (Random Forest Regressor), Линейная корреляция (f\_regression)
14. Случайное Лассо (RandomizedLasso), Сокращение признаков Случайными деревьями (Random Forest Regressor), Линейная корреляция (f\_regression)
15. Случайное Лассо (RandomizedLasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Линейная корреляция (f\_regression)
16. Линейная регрессия (LinearRegression), Случайное Лассо (RandomizedLasso), Линейная корреляция (f\_regression)
17. Гребневая регрессия (Ridge), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Линейная корреляция (f\_regression)
18. Лассо (Lasso), Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE), Линейная корреляция (f\_regression)
19. Линейная регрессия (LinearRegression), Гребневая регрессия (Ridge), Лассо (Lasso), Случайное Лассо (RandomizedLasso)
20. Линейная регрессия (LinearRegression), Гребневая регрессия (Ridge), Лассо (Lasso), Линейная корреляция (f\_regression).

Контрольные вопросы по лабораторной работе:

1. В чем суть задачи ранжирования признаков?

2. Расскажите подробнее об используемых вами моделях.
3. В чем суть регрессионной проблемы Фридмана?

### **Лабораторная работа 3. Деревья решений**

Часть 1. По данным о пассажирах Титаника решите задачу классификации (с помощью дерева решений), в которой по различным характеристикам пассажиров требуется найти у выживших пассажиров два наиболее важных признака из трех рассматриваемых (по варианту). Пример решения задачи можно посмотреть здесь: [1] (стр.188). Скачать данные можно по ссылке: <https://www.kaggle.com/datasets/heptapod/titanic>

- |                          |                         |
|--------------------------|-------------------------|
| 1 Pclass,Name,Sex        | 2 Name,Sex,Age          |
| 3 Sex,Age,SibSp          | 4 Age,SibSp,Parch       |
| 5 SibSp,Parch,Ticket     | 6 Parch,Ticket,Fare     |
| 7 Ticket,Fare,Cabin      | 8 Fare,Cabin,Embarked   |
| 9 Pclass, Cabin,Embarked | 10 Name, Cabin,Embarked |
| 11 Sex,Cabin,Embarked    | 12 Age,Cabin,Embarked   |
| 13 SibSp,Cabin,Embarked  | 14 Parch,Cabin,Embarked |
| 15 Ticket,Cabin,Embarked | 16 SibSp,Parch,Ticket   |
| 17 Age,Fare, ,Embarked   | 18 Pclass, Age,Ticket,  |
| 19 Name, Fare, Embarked  | 20 Pclass, Parch, Fare  |

Часть 2. Решите с помощью библиотечной реализации дерева решений задачу из лабораторной работы «Веб-сервис «Дерево решений» по предмету «Методы искусственного интеллекта» на 99% ваших данных. Проверьте работу модели на оставшемся проценте, сделайте вывод.

#### Контрольные вопросы по лабораторной работе:

1. Расскажите подробнее об используемой вами модели дерева решений.
2. Какая типовая задача машинного обучения решается в обеих частях лабораторной?

3. Какими еще моделями можно решить поставленные задачи?

### Лабораторная работа 4. Кластеризация

Использовать метод кластеризации по варианту для данных из таблицы 1 по варианту (таблица 9), самостоятельно сформулировав задачу. Интерпретировать результаты и оценить, насколько хорошо он подходит для решения сформулированной вами задачи.

Таблица 9. Варианты

Остаток от деления номера варианта на 5	Алгоритм кластеризации
0	K-means
1	dendrogram
2	linkage
3	t-SNE
4	DBSCAN

Контрольные вопросы по лабораторной работе:

1. Какую задачу для решения вы сформулировали?
2. Как вы оценили качество кластеризации?
3. Расскажите подробнее об используемом вами алгоритме кластеризации.

### Лабораторная работа 5. Регрессия

Использовать регрессию по варианту для данных из таблицы 1 по варианту (таблица 10), самостоятельно сформулировав задачу. Оценить, насколько хорошо она подходит для решения сформулированной вами задачи.

Таблица 10. Варианты

Остаток от деления номера варианта на 5	Тип регрессии
0	Линейная регрессия



	1	Полиномиальная регрессия
	2	Логистическая регрессия
	3	Лассо-регрессия
	4	Гребневая регрессия

Контрольные вопросы по лабораторной работе:

1. Какую задачу для решения вы сформулировали?
2. Как вы оценили качество решения?
3. Расскажите подробнее об используемой вами модели.

### **Лабораторная работа 6. Нейронная сеть**

Использовать нейронную сеть (четные варианты – MLPRegressor, нечетные – MLPClassifier) для данных из таблицы 1 по варианту, самостоятельно сформулировав задачу. Интерпретировать результаты и оценить, насколько хорошо она подходит для решения сформулированной вами задачи.

Контрольные вопросы по лабораторной работе:

1. Какую задачу для решения вы сформулировали?
2. Как вы оценили качество решения?
3. Расскажите подробнее об используемой вами модели.

### **Лабораторная работа 7. Рекуррентная нейронная сеть и задача генерации текста**

Выбрать художественный текст (четные варианты – русскоязычный, нечетные – англоязычный) и обучить на нем рекуррентную нейронную сеть для решения задачи генерации. Подобрать архитектуру и параметры так, чтобы приблизиться к максимально осмысленному результату. Далее разбиться на пары четный-нечетный вариант, обменяться разработанными сетями и проверить, как архитектура товарища справляется с вашим текстом.

В завершении подобрать компромиссную архитектуру, справляющуюся достаточно хорошо с обоими видами текстов.

Контрольные вопросы по лабораторной работе:

1. Какой текст вы выбрали и почему?
2. Какова архитектура вашей сети?
3. На скольких эпохах вы ее обучали и каких результатов удалось добиться?

## Список литературы

1. Теория и практика машинного обучения : учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. – Ульяновск : УлГТУ, 2017. – 290 с. URL: <http://venec.ulstu.ru/lib/disk/2017/191.pdf> . Дата обращения 11.08.2022.
2. [Электронный ресурс]: Справочник по языку Python. URL: <https://pythonworld.ru/samouchitel-python>. Дата обращения 11.08.2022.

Учебное издание

Методы искусственного интеллекта и машинное обучение

Практикум

Автор-составитель Воронина Валерия Вадимовна

Редактор Н. А. Евдокимова

Подписано в печать 26.11.2020.

Усл. печ. л. 2,79. Заказ

Объем 419 Кб.

Ульяновский государственный технический университет

432027, г. Ульяновск, ул. Сев. Венец, д. 32.

ИПК «Венец» УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32