

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Смоленский государственный университет»

Кафедра информатики

*«Утверждаю»*

Проректор по учебно-  
методической работе  
\_\_\_\_\_ Устименко Ю.А.  
«2» сентября 2021 г.

**Рабочая программа дисциплины**  
**Б1.В.ДВ.04.02 РАЗРАБОТКА АНАЛИТИЧЕСКИХ СИСТЕМ**

Направление подготовки: **01.04.02 Прикладная математика и информатика**

Направленность (профиль): **Прикладные интернет-технологии**

Форма обучения: очная

Курс – 2

Семестр – 3

Всего зачетных единиц – 2, часов – 72

Форма отчетности: зачет – 3 семестр

Программу разработал  
кандидат технических наук, доцент Т.А. Самойлова

Одобрена на заседании кафедры  
«26» августа 2021 г., протокол № 1

Смоленск  
2021

## 1. Место дисциплины в структуре ОП

Дисциплина «Разработка аналитических систем» относится к дисциплинам по выбору и является вспомогательной для производственной практики магистров на предприятиях. Она изучается в 3 семестре.

При изучении данной дисциплины необходимы компетенции студентов, сформированные при изучении таких дисциплин, как «Информационные системы», «Системы искусственного интеллекта», «Теория вероятностей и математическая статистика». Она знакомит магистра с новыми подходами к разработке интеллектуальных приложений, предназначенных для решения проблем современных предприятий, обусловленных ростом объема данных, необходимостью интеграции с партнёрами и новыми подразделениями, а также внедрения средств анализа данных. Дисциплина «Разработка аналитических систем» является обеспечивающей для работы магистров на предприятиях. Дисциплина способствует формированию у магистра целостного представления о возможностях разработки сетевых информационных систем в ходе решения задач интеллектуальной обработки данных и прикладных математических задач. Приобретенные знания помогут магистру выбрать направление будущих научных исследований.

Изучение курса основано на традиционных методах высшей школы, тесной взаимосвязи со смежными курсами, а также на использовании современного программного обеспечения для машинного обучения.

## 2. Планируемые результаты обучения по дисциплине

Компетенция	Индикаторы достижения <i>(в соответствии с разделом 7 общей характеристики ОП ВО)</i>
<b>ПК-1.</b> Способен осуществлять поиск, анализ, систематизацию научной информации в области прикладной математики и информатики для реализации научно-исследовательских проектов и решения прикладных задач.	<b>Знает:</b> теоретические основы и технологии организации научно-исследовательской деятельности. <b>Умеет:</b> осуществлять поиск, анализ, систематизацию научной (в том числе юридической) информации в области прикладной математики и информатики для реализации научно-исследовательских проектов и решения прикладных задач, подготавливать и представлять для обсуждения научно-исследовательские работы. <b>Владеет:</b> навыками организации и проведения научно-исследовательской деятельности в ходе выполнения профессиональных функций.
<b>ПК-2.</b> Способен проектировать программное и информационное обеспечение компьютерных сетей, вычислительные модели и модели данных для реализации элементов новых (или известных) программных продуктов.	<b>Знает:</b> структуру программного обеспечения, основные требования к его проектированию, современные языки программирования, технологии программирования, методики разработки и анализа блок-схем алгоритмов. <b>Умеет:</b> проектировать программное и информационное обеспечение компьютерных сетей, выявлять требования к программным продуктам, создавать блок-схемы алгоритмов функционирования разрабатываемых программных продуктов, оценивать их вычислительную сложность. <b>Владеет:</b> методами и приемами формализации и алгоритмизации поставленных проектных задач по созданию программного обеспечения.

<p><b>ПК-3.</b> Способен разрабатывать программное обеспечение, в том числе драйверы устройств, компиляторы, загрузчики, сборщики, системные утилиты.</p>	<p><b>Знает:</b> архитектуру аппаратной платформы, синтаксис, принципы и особенности программирования (в том числе кроссплатформенного, распределенного и параллельного программирования), стандартные библиотеки выбранного языка программирования, стандарты реализации интерфейсов устройств, технологии разработки и отладки программных продуктов, принципы информационного построения сетевого взаимодействия, методики тестирования программного обеспечения.</p> <p><b>Умеет:</b> использовать методы и приемы формализации и алгоритмизации поставленных задач, применять выбранный язык программирования для написания программного кода, осуществлять отладку программного обеспечения.</p> <p><b>Владеет:</b> основными методами и приемами разработки и отладки программных продуктов.</p>
---	---

### 3. Содержание дисциплины

В дисциплине «Разработка аналитических систем» рассматриваются следующие темы.

**1. Средства Python для работы с алгоритмами анализа данных.** Необходимые сведения из теории вероятностей и математической статистики Типы данных. Математическое ожидание, медиана, мода, стандартное отклонение, дисперсия. Функции распределения, плотности вероятностей. Ковариация и корреляция. Условная вероятность. Методы и атрибуты объектов pandas. Теорема Байеса. Байесовские методы. Кластеризация. Измерение энтропии. Деревья решений. Ensemble learning. SVM. Метод K-ближайших соседей. Уменьшение размерности. Метод анализа главных компонент. ROC-анализ.

**2. Библиотека Scikit-learn для работы с алгоритмами анализа данных.** NumPy как средство эффективной и SciPy как средство интеллектуальной обработки данных. Предварительная обработка и очистка данных. Парциальная предобработка. Спектральная обработка. Корреляционный анализ. Выявление дубликатов и противоречий Выбор подходящей модели и обучающего алгоритма. Построение модели классификации. Признаки и подготовка признаков. Классификация с помощью Scikit-learn. Бинарная и многоклассовая классификация. Оценка качества классификации.

**3. Поиск ассоциативных правил в анализе данных.** Понятия и методы выявления закономерностей алгоритмами ассоциаций. Использование выявленных закономерностей в для предсказания неизвестных значений. Поддержка, достоверность, лифт. Алгоритмы Apriori, FP-Growth, Eclat. Анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях. Функция ошибок. Функционал качества.

**4. Сверточные нейронные сети (СНС).** Свертка. Свертка функций. Свертка изображений. Сверточные вычисления. Биологический прототип СНС. Задачи СНС. Архитектура СНС: блок выделения признаков и блок принятий решений. Сверточный слой. Одноканальная и многоканальная свертки. Ядро, фильтр, коллекция фильтров. Функции активации. Пулинг. Batch-нормализация. Полносвязный слой. Обратное распространение ошибки back-propagation. Модификации СНС. Использование предобученных нейронных сетей для решения задачи классификации изображений (Transfer Learning). Расширение обучающего датасета с использованием техники аугментации изображений (Image Augmentation).

5. **Обучение с подкреплением в аналитических системах.** Постановка задачи. Взаимодействие агента с окружающей средой. Оценка поведения, поощрения окружающей среды. Подкрепление и значение состояния. Агенты с одним состоянием. Действие с наилучшей ожидаемой прибылью: жадный алгоритм. Решения методом динамического программирования. Байесовский подход к многоруким бандитам.

#### 4. Тематический план

№ п/п	Разделы и темы	Всего часов	Формы занятий			
			лекции	практические занятия	лабораторные занятия	самостоятельная работа
1	Средства Python для работы с алгоритмами анализа данных.	18	4	–	4	10
2	Библиотека Scikit-learn для работы с алгоритмами анализа данных.	18	4	–	4	10
3	Поиск ассоциативных правил в анализе данных.	12	2	–	2	8
4	Сверточные нейронные сети.	12	2	–	2	8
5	Обучение с подкреплением в аналитических системах.	12	2	–	2	8
ИТОГО		72	14	–	14	44

#### 5. Виды образовательной деятельности

##### Занятия лекционного типа

**1. Средства Python для работы с алгоритмами анализа данных.** Необходимые сведения из теории вероятностей и математической статистики Типы данных. Математическое ожидание, медиана, мода, стандартное отклонение, дисперсия. Функции распределения, плотности вероятностей. Ковариация и корреляция. Условная вероятность. Методы и атрибуты объектов pandas. Теорема Байеса. Байесовские методы. Кластеризация. Измерение энтропии. Деревья решений. Ensemble learning. SVM. Метод К-ближайших соседей. Уменьшение размерности. Метод анализа главных компонент. ROC-анализ.

**2. Библиотека Scikit-learn для работы с алгоритмами анализа данных.** NumPy как средство эффективной и SciPy как средство интеллектуальной обработки данных. Предварительная обработка и очистка данных. Парциальная предобработка. Спектральная обработка. Корреляционный анализ. Выявление дубликатов и противоречий Выбор подходящей модели и обучающего алгоритма. Построение модели классификации. Признаки и подготовка признаков. Классификация с помощью Scikit-learn. Бинарная и многоклассовая классификация. Оценка качества классификации.

**3. Поиск ассоциативных правил в анализе данных.** Понятия и методы выявления закономерностей алгоритмами ассоциаций. Использование выявленных закономерностей в для предсказания неизвестных значений. Поддержка, достоверность, лифт. Алгоритмы Apriori, FP-Growth, Eclat. Анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях. Функция ошибок. Функционал качества.

**4. Сверточные нейронные сети (СНС).** Свертка. Свертка функций. Свертка изображений. Сверточные вычисления. Биологический прототип СНС. Задачи СНС. Архитектура СНС: блок выделения признаков и блок принятий решений. Сверточный слой.

Одноканальная и многоканальная свертки. Ядро, фильтр, коллекция фильтров. Функции активации. Пулинг. Batch-нормализация. Полносвязный слой. Обратное распространение ошибки back-propagation. Модификации СНС. Использование предобученных нейронных сетей для решения задачи классификации изображений (Transfer Learning). Расширение обучающего датасета с использованием техники аугментации изображений (Image Augmentation).

**5. Обучение с подкреплением в аналитических системах.** Постановка задачи. Взаимодействие агента с окружающей средой. Оценка поведения, поощрения окружающей среды. Подкрепление и значение состояния. Агенты с одним состоянием. Действие с наилучшей ожидаемой прибылью: жадный алгоритм. Решения методом динамического программирования. Байесовский подход к многоруким бандитам.

### **Занятия семинарского типа**

Не предусмотрены

### **Лабораторные работы**

№1-2. Инструменты анализа данных в Python. Библиотека Pandas.

№3-4. Библиотеки Python для машинного обучения.

№5. Создание Python - приложений с использованием моделей поиска ассоциативных правил.

№6. Свёрточная нейронная сеть для распознавания изображений.

№7. Машинное обучение с подкреплением в аналитических системах.

Задания для лабораторных работ, размещены в системе дистанционного обучения СмолГУ ([www.moodle.smolgu.ru](http://www.moodle.smolgu.ru)). На занятиях для каждой работы задание предоставляется студентам в электронном виде.

### **Самостоятельная работа**

Текущая самостоятельная работа студента направлена на углубление и закрепление знаний студентов, развитие практических умений. Она заключается в работе с лекционными материалами, поиске и обзоре литературы и электронных источников, информации по заданным темам курса, опережающей самостоятельной работе, в изучении тем, вынесенных на самостоятельную проработку, подготовке к лабораторным занятиям.

Самостоятельная внеаудиторная работа студентов включает:

- проработку лекционного материала, составление конспекта лекций по темам, вынесенным на самостоятельное изучение;
- выполнение домашних заданий;
- подготовку к защите лабораторных работ.

### **Темы для самостоятельного изучения**

1. Среды разработки аналитических систем.
2. Библиотеки для разработки аналитических систем в языках `c#` и `python`.
3. Пакеты программ машинного обучения.
4. Алгоритмы поиска ассоциативных правил.
5. Применение нейронных сетей в аналитических системах.
6. Обучение с подкреплением. Взаимодействие агента с окружающей средой.
7. Оценка качества обучения с подкреплением в аналитических системах.

### **Учебно-методическое обеспечение для самостоятельной работы**

1. Миркин Б. Г. Введение в анализ данных. — М.: Юрайт. 2020. 175 с.

## **6. Критерии оценивания результатов освоения дисциплины (модуля)**

### **6.1. Оценочные средства и критерии оценивания для текущей аттестации**

#### **Теоретические вопросы**

1. Средства Python для работы с алгоритмами анализа данных.

2. Необходимые сведения из теории вероятностей и математической статистики Типы данных. Математическое ожидание, медиана, мода, стандартное отклонение, дисперсия.
3. Функции распределения, плотности вероятностей. Ковариация и корреляция. Условная вероятность.
4. Методы и атрибуты объектов pandas. Теорема Байеса. Байесовские методы. Кластеризация. Измерение энтропии.
5. Деревья решений. Ensemble learning. SVM. Метод K-ближайших соседей. Уменьшение размерности. Метод анализа главных компонент. ROC-анализ.
6. Библиотека Scikit-learn для работы с алгоритмами МО. NumPy как средство эффективной и SciPy как средство интеллектуальной обработки данных.
7. Предварительная обработка и очистка данных. Парциальная предобработка. Спектральная обработка. Корреляционный анализ. Выявление дубликатов и противоречий. Выбор подходящей модели и обучающего алгоритма.
8. Построение модели классификации. Признаки и подготовка признаков. Классификация с помощью Scikit-learn. Бинарная и многоклассовая классификация. Оценка качества классификации.
9. Поиск ассоциативных правил в анализе данных. Понятия и методы выявления закономерностей алгоритмами ассоциаций.
10. Использование выявленных закономерностей для предсказания неизвестных значений. Поддержка, достоверность, лифт.
11. Алгоритмы Apriori, FP-Growth, Eclat. Анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях. Функция ошибок. Функционал качества.
12. Сверточные нейронные сети (СНС). Свертка. Свертка функций. Свертка изображений. Сверточные вычисления.
13. Биологический прототип СНС. Задачи СНС. Архитектура СНС: блок выделения признаков и блок принятий решений. Сверточный слой.
14. Одноканальная и многоканальная свертки. Ядро, фильтр, коллекция фильтров. Функции активации.
15. Пулинг. Batch-нормализация. Полносвязный слой. Обратное распространение ошибки back-propagation. Модификации СНС.
16. Использование предобученных нейронных сетей для решения задачи классификации изображений (Transfer Learning).
17. Расширение обучающего датасета с использованием техники аугментации изображений (Image Augmentation).
18. Обучение с подкреплением в аналитических системах. Постановка задачи. Взаимодействие агента с окружающей средой. Оценка поведения, поощрения окружающей среды. Подкрепление и значение состояния.
19. Агенты с одним состоянием. Действие с наилучшей ожидаемой прибылью: жадный алгоритм.
20. Решения методом динамического программирования. Байесовский подход к многоруким бандитам.

### Критерии оценивания теоретических вопросов

#### 1. Нормы оценивания ответов на теоретические вопросы

№ п/п	Теоретический вопрос	Количество баллов (*)
1	Дан краткий ответ на поставленный вопрос	1 балл
2	Дан развернутый ответ на вопрос с анализом результатов	2 балла

(\*) Возможна градация в 0,25 балла.

2. Шкала оценивания. Оценка «зачтено» за ответы на теоретические вопросы выставляется, если набрано не менее 3 баллов при ответе на три вопроса, в противном случае выставляется «не зачтено».

### **Задания для лабораторных занятий**

Задачи по темам курса предложены к каждому лабораторному занятию.

#### **Образцы заданий**

##### **№1-2. Инструменты анализа данных в Python. Библиотека Pandas.**

**Задание 1.** Используя Pandas, выполнить импорт набора данных \*.CSV, содержащий пропущенные данные. Выполнить:

- замену отсутствующих значений нулями
- замену отсутствующих средними по столбцу
- удаление строк с отсутствующими элементами

**Задание 2.** Используя Pandas, выполнить импорт набора данных \*.CSV, содержащих строковые и числовые поля. Выполнить (с выводом таблиц):

- Отбор колонок (2-4)
- Отбор строк в некотором диапазоне
- Фильтрацию данных по числовому значению (>,<) и по заданному значению строкового поля.
- Подсчет количеств строк с каждым значением строкового поля

##### **№3-4. Библиотеки Python для машинного обучения.**

**Задание 1.** Методом деревьев решений выполните обучение модели классификации на конкретных индивидуальных данных. Оцените качество модели на тестовых данных, эквивалентных обучающим. Результаты: оценки качества - точность, полнота, f-мера. Постройте график зависимости класса от двух входных параметров (например, возраст, доход). До обучения модели, постройте корреляционную матрицу параметров, которая показывает, какие параметры будут влиять на результат. Оставьте для обучения 3-4 наиболее значимых параметра. Замените категориальные параметры числовыми. Тестовую выборку сформируйте методом кросс - валидации. Подберите параметры метода, обеспечивающие лучшее качество.

**Задание 2.** Обучите модель методом деревьев решений по вашим данным, используя алгоритмы C4.5 и CART. Сравните результаты

**Задание 3.** Обучите модель методом опорных векторов по вашим данным. Выборка ваша из 3-7 классов. До обучения модели, постройте корреляционную матрицу параметров, которая показывает, как параметры будут влиять на результат. Тестовую выборку сформируйте методом кросс - валидации. Постройте матрицу неточности. Выполните ROC-анализ. Вычислите AUC модели.

**Задание 3.** Разработайте программы регрессии методом опорных векторов с использованием линейных, полиномиальных и RBF-ядер для индивидуальных данных (UCI Machine Learning Repository - <http://archive.ics.uci.edu/ml/>). Вычислите оценки качества для тестовой и обучающей выборки: среднеквадратичную ошибку регрессии - mean\_squared\_error (MSE), коэффициент детерминации  $r^2$  score. Сравните результаты для разных типов ядер. Постройте графики зависимости целевой функции от признаков выборки.

##### **№5. Создание Python - приложений с использованием моделей поиска ассоциативных правил.**

**Задание 1.** Средствами Python создайте приложение для решения задачи поиска ассоциативных правил с помощью алгоритма Apriori. В качестве предметной области взять розничную торговлю:

- определение товаров, которые стоит продвигать совместно;
- выбор местоположения товара в магазине;

- анализ потребительской корзины; прогнозирование спроса.

### №6. Сверточная нейронная сеть для распознавания изображений.

**Задание 1.** Разработать программу для обучения сверточной нейронной сети распознаванию рукописных цифр из библиотеки MNIST. Сеть обязательно имеет входной, сверточный, пулинговый и выходной слой. Сохраните обученную модель на диск.

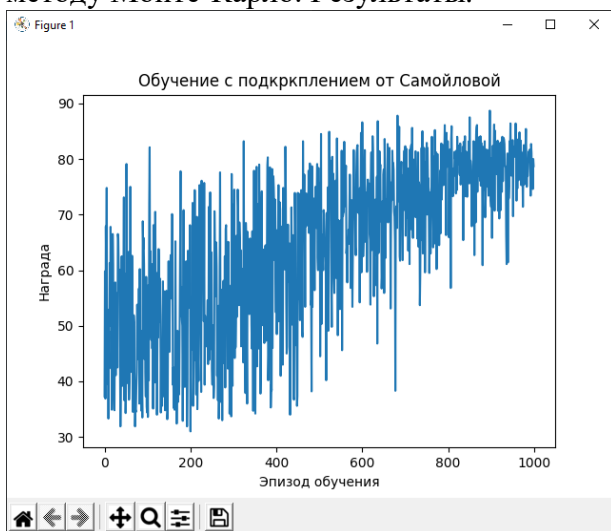
**Задание 2.** Добавьте в модель еще один сверточный слой (в паре с пулинговым). Добавьте простой скрытый слой. Сравните точности моделей. Сравните точность с моделью DNNполносвязной сети (для MNIST).

**Задание 3.** Разработать программу, которая по обученной и сохраненной модели (задание 1) распознает рукописные цифры. Распознайте 1) 3-4 изображения из библиотеки MNIST; 2) нарисованное вами изображение рукописной цифры размера 28\*28.

**Задание 4.** По готовой обученной модели сврточной нейронной сети (CNN) VGG16 распознайте 2-3 произвольных изображения (самолеты, морские суда, автомобили, любые животные и т.п.). Организуйте вывод изображения, используя cv2.

### №7. Машинное обучение с подкреплением в аналитических системах.

**Задание 1.** Создайте Python - приложение в тренажерном зале OpenAI для выполнения обучения с подкреплением в среде «MountainCar», когда взаимодействия агента с окружающей средой выполняется по алгоритму градиента политики. Тренировочный процесс проходит по методу Монте-Карло. Результаты:



### Критерии оценивания выполнения лабораторных работ

1. Нормы оценивания каждой лабораторной работы:

№п/п	Структурная часть работы	Количество баллов (*)
1	Ответ на теоретические вопросы по теме лабораторной работы	1 балл
2	Демонстрация выполнения конкретного задания, предложенного для самостоятельного решения на лабораторной работе	2 балла

(\*) с возможностью градации до 0,25 балла.

2. Шкала оценивания. Оценка «зачтено» за лабораторную работу выставляется, если набрано не менее 2 баллов, в противном случае за работу выставляется «не зачтено».

### 6.2. Оценочные средства и критерии оценивания для промежуточной аттестации

#### Зачетная контрольная работа

1. Разработайте приложение, которое, используя обученную вами модель сверточной нейронной сети для распознавания изображения, определит, что изображено на



- созданном вами рисунке (одежда, номер автомобиля, зверь). Испытайте проект для разных рисунков.
- Разработайте приложение, которое выполняет импорт набора данных \*.CSV, содержащий пропущенные данные. Заполните данные, используя библиотечные средства Python.

### Критерии оценивания зачетной контрольной работы

#### 1. Нормы оценивания работы

№ п/п	Структурная часть контрольной работы	Количество баллов (*)
1	Правильно реализован каждый метод решения	1 балл
2	Анализ результатов	2 балла

(\*) Возможна градация в 0,25 балла.

#### 2. Шкала оценивания работы:

п/п	Оценка	Количество баллов
1	Отлично	4,75-5
2	Хорошо	3,75-4,5
3	Удовлетворительно	3-3,5
4	Неудовлетворительно	менее 3

### Критерий получения зачета

Зачет выставляется по результатам работы студента в течение семестра согласно Положению о текущем контроле успеваемости и промежуточной аттестации обучающихся в федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Смоленский государственный университет» (утверждено приказом ректора № 01-113 от 26.09.2019 г.; внесены дополнения приказом ректора № 01-48 от 30.04.2020).

Для получения зачета студент должен:

- уметь отвечать на теоретические вопросы, рассмотренные на лекциях;
- уметь решать задачи, предложенные на лабораторных занятиях;
- уметь решать задачи, предложенные на зачетной контрольной работе.

## 7. Перечень основной и дополнительной учебной литературы

### 7.1. Основная литература

- Кудрявцев В. Б. Интеллектуальные системы: учебник и практикум для вузов – М.: Юрайт, 2021. 165 с. Текст: электронный // Образовательная платформа Юрайт [сайт]. – URL: <https://urait.ru/bcode/423761>
- Лаврищева Е. М. Программная инженерия и технологии программирования сложных систем: учебник для вузов / Е. М. Лаврищева. – 2-е изд., испр. и доп. – Москва: Издательство Юрайт, 2021. – 432 с. – (Высшее образование). – ISBN 978-5-534-07604-2. – Текст: электронный // Образовательная платформа Юрайт [сайт]. – URL: <https://urait.ru/bcode/470923>
- Сысолетин Е. Г. Ростунцев С. Д. Разработка интернет - сервисов: учебное пособие для вузов / под научной редакцией Л. Г. Доросинского. – Москва: Издательство Юрайт, 2021. – 90 с. – (Университеты России). – ISBN 978-5-9916-9975-4. – Текст: электронный // ЭБС Юрайт [сайт]. – URL: <https://urait.ru/bcode/398246> .

### 7.2. Дополнительная литература

- Бенджио И., Гудфеллоу Я., Курвилль А. Глубокое обучение. – М.: ДМК Пресс. 2021. 652 с.
- Миркин Б. Г. Введение в анализ данных. – М.: Юрайт. 2021. 175 с.
- Мунерман В.И., Самойлова Т.А. Обучение методам разработки информационно-аналитических систем на основе облачных технологий. Системы высокой доступности, 4, 2016, т.12, с. 3-11
- Фальк К. Рекомендательные системы на практике. М.: ДМК Пресс. 2020. 448с. 7.3.

### 7.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Система дистанционного обучения СмолГУ (moodle.smolgu.ru).
2. Рекомендательные системы (видеолекция). А.Просветов, К.Коточигов. <https://www.youtube.com/watch?v=QAoL6cptzC0>
3. <http://www.niisi.ru/iont/ni> - Российская ассоциация нейроинформатики

### 8. Материально-техническое обеспечение

Для занятий необходимы:

1. проектор;
2. интерактивная доска;
3. персональные компьютеры.

Для самостоятельной работы подготовлены аудитории № 224, 226, 230, 234 с выходом в Интернет, оснащенные компьютерами IBMPC с процессорами IntelCore 7 и оперативной памятью не менее 16 ГБ.

### 9. Программное обеспечение

1. Python 3.9.5; среда разработки приложений PyCharm

ДОКУМЕНТ ПОДПИСАН  
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 6314D932A1EC8352F4BBFDEFD0AA3F30

Владелец: Артеменков Михаил Николаевич

Действителен: с 21.09.2022 до 15.12.2023