ООО «ЭЛЬБРУС БУТКЕМП»

УТВЕРЖДАЮ
Генеральный директор
_____ / Бабаян Г. Г. /
«19» октября 2026 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА

Профессиональная переподготовка

Специалист по анализу данных Data Scientist

560 академических часов
14 недель · 70 учебных дней

Форма обучения: очно-заочная.
Реализация с применением электронного обучения и дистанционных образовательных технологий.

Программа применяется к новым наборам после её утверждения. Условия ранее заключённых договоров настоящей редакцией не изменяются.

Москва 2026

2. ОПИСЬ КОМПЛЕКТА ДОКУМЕНТОВ

Раздел	Содержание
1	Титульный лист
2	Опись комплекта документов
3	Общие положения
4	Цель и задачи программы
5	Планируемые результаты обучения
6	Учебный план
7	Календарный учебный график
8	Рабочая программа и текущая аттестация
9	Организационно-педагогические условия
10	Программа итоговой аттестации
11	Оценочные средства и критерии итоговой аттестации
12	Кадровое обеспечение

Учебный план и календарный график являются неотъемлемыми частями настоящей программы и утверждаются вместе с ней.

3. ОБЩИЕ ПОЛОЖЕНИЯ

3.1. Характеристика программы

Дополнительная профессиональная программа профессиональной переподготовки «Специалист по анализу данных Data Scientist» трудоёмкостью 560 академических часов (далее - Программа) разработана в соответствии с:

• Федеральным законом от 29.12.2012 № 273-ФЗ «Об образовании в Российской Федерации»;

• приказом Минобрнауки России от 24.03.2025 № 266 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;

• постановлением Правительства РФ от 11.10.2023 № 1678 об утверждении Правил применения электронного обучения и дистанционных образовательных технологий, в действующей редакции;

• профессиональным стандартом «Специалист по большим данным», утверждённым приказом Минтруда России от 06.07.2020 № 405н.

При определении компетенций учитываются требования ФГОС высшего образования по направлению 09.03.02 «Информационные системы и технологии» (приказ от 19.09.2017 № 926, в действующей редакции). Программа реализуется в дополнительном профессиональном образовании.

3.2. Характеристика профессиональной деятельности выпускников:

– **область профессиональной деятельности** в соответствии с ФГОС в сферах: связь, информационные и коммуникационные технологии (в сфере исследования, разработки, внедрения и сопровождения информационных технологий и систем).

– **основная цель вида профессиональной деятельности:** Создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг

– **трудовые функции**:

Обобщенные трудовые функции			Трудовые функции		
код	наименование	уровень квали-фикации	наименование	код	уровень (подуровень) квалификации

A	Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	6	Выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных.	A/01.6	6
			Планирование и организация аналитических работ с использованием технологий больших данных.	A/02.6	6
			Подготовка данных для проведения аналитических работ по исследованию больших данных.	A/03.6	6
			Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика.	A/04.6	6
B	Управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации	7	Анализ потребности заинтересованных лиц и/или подразделений организации в исследовании больших данных	B/01.7	7
			Управление получением, хранением, передачей, обработкой больших данных	B/05.7	7
D	Разработка и внедрение новых методов и технологий исследования больших данных	8	Совершенствование и разработка новых методов, моделей, алгоритмов и инструментальных средств.	D/01.8	8
			Проведение испытаний и разработка рекомендаций по внедрению новых методов, моделей и технологий.	D/02.8	8

– **вид программы**: практико-ориентированная.

3.3. Категория обучающихся по:

– основной специальности: Аналитик данных, Data Scientist,

– смежным специальностям: Дата-инженер, маркетолог, бизнес-стратег .

К освоению Программы допускаются лица, имеющие или получающие среднее профессиональное и (или) высшее образование, а также имеющие базовые знания

синтаксиса Python.

3.4. Актуальность программы: Концепция анализа данных приобретает всё большую значимость для специалистов различных областей, поскольку в современном мире практически каждая сфера деятельности активно использует данные для принятия решений. Понимание методов сбора, обработки и анализа данных позволяет не только улучшить операционные процессы и повысить эффективность работы, но и выявлять скрытые закономерности, оптимизировать ресурсы и прогнозировать будущее развитие проектов. Для слушателей курсов переподготовки эти навыки открывают новые возможности карьерного роста, особенно в условиях быстро меняющегося рынка труда, где работодатели всё чаще требуют умения работать с большими массивами информации.

Актуальность изучения анализа данных также обусловлена развитием технологий и цифровой трансформацией. Специалисты, обладающие компетенциями в этой области, становятся более востребованными на рынке, поскольку могут использовать современные инструменты и платформы для решения практических задач. Это не только повышает их конкурентоспособность, но и даёт возможность стать активными участниками инновационных процессов в компаниях, будь то малый бизнес, государственные структуры или крупные корпорации.

3.5. Объём программы (трудоёмкость)

Общая трудоёмкость составляет 560 академических часов. Срок освоения - 14 недель, 70 учебных дней. Контактная работа с педагогическими работниками с применением электронного обучения и дистанционных образовательных технологий составляет 354 часа: 98 часов лекций, 252 часа практических занятий и 4 часа итоговой аттестации. Самостоятельная работа составляет 206 часов.

3.6. Документ, выдаваемый после завершения обучения

При успешном освоении программы и прохождении итоговой аттестации выдаётся диплом о профессиональной переподготовке образца, установленного организацией. Лицам, обучающимся параллельно по программам СПО или ВО, диплом выдаётся одновременно с получением соответствующего документа об образовании и квалификации. Лицам, не прошедшим итоговую аттестацию или освоившим часть программы, выдаётся справка об обучении или о периоде обучения установленного организацией образца.

4. ЦЕЛЬ И ЗАДАЧИ ПРОГРАММЫ

4.1. Цель программы заключается в получении компетенций, необходимых для профессиональной деятельности в сфере проектирования, разработки и интеграция информационных ресурсов в локальной сети и информационно-телекоммуникационной сети «Интернет».

4.2. Задачи

В процессе обучения решаются следующие задачи:

1. Освоение основ Python и среды разработки: слушатели должны научиться настраивать рабочее окружение для программирования и освоить базовые техники разработки на Python, что позволит эффективно работать с данными и разрабатывать свои первые проекты (pet-проекты).
2. Применение методов описательной статистики для анализа данных: обучение основам сбора, обработки и визуализации данных с помощью Python, а также базовым статистическим методам для описания и интерпретации полученных данных.
3. Изучение основ теории вероятностей и статистики для анализа данных и проведения А/Б тестирования: слушатели приобретут навыки работы с вероятностными моделями и статистическими методами, что даст возможность грамотно оценивать гипотезы и проводить эксперименты, например, А/Б тесты.
4. Применение линейной алгебры и анализа для решения задач машинного обучения: участники научатся использовать математические основы для построения моделей машинного обучения с учителем, включая регрессию, классификацию и работу с временными рядами.
5. Изучение методов кластеризации и временных рядов: слушатели освоят методы безучительного обучения и смогут анализировать данные, распределённые по временным интервалам.
6. Изучение нейросетевого моделирования для решения сложных задач: участники познакомятся с основами построения нейронных сетей, будут работать с рекомендательными системами и мультимодальными задачами.
7. Применение SQL и PySpark для работы с большими данными: слушатели приобретут навыки работы с большими массивами данных с использованием SQL и PySpark, что даст возможность интегрировать и анализировать данные в средах с большими объёмами информации.
8. Интеграция полученных знаний через продакшен собственного проекта: участники курса смогут применить полученные знания на практике, разработав и завершив свой pet-проект, который продемонстрирует их компетенции в сфере анализа данных и машинного обучения.

5. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

5.1. Программа направлена совершенствование:

– универсальных компетенций:

- Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах) (УК-4);
- Способен воспринимать межкультурное разнообразие общества в социально-историческом, этическом и философском контекстах (УК-5);

– общепрофессиональных компетенций:

- Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности (ОПК-1);
- Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности (ОПК-2);
- Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности (ОПК-3);
- Способен участвовать в разработке технической документации, связанной с профессиональной деятельностью с использованием стандартов, норм и правил (ОПК-4).
- Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем (ОПК-8)

– профессиональных компетенций::

производственно-технологическая деятельность

- Способен выявлять, формировать и согласовывать требования к результатам аналитических работ с применением технологий больших данных (ПК-1);
- Способен подготовить данные для проведения аналитических работ по исследованию больших данных, включая их получение и очистку (ПК-2).

организационно-управленческая деятельность:

- Способен планировать и организовывать аналитические работы с использованием технологий больших данных, включая разработку расписания и распределение задач (ПК-3);
- Способен анализировать потребности заинтересованных лиц и подразделений в исследовании больших данных и формировать предложения по их удовлетворению (ПК-4);

сервисно-эксплуатационная деятельность:

- Способен управлять процессами получения, хранения, передачи и обработки больших данных, включая мониторинг и поддержку инфраструктуры (ПК-5);
- Способен совершенствовать и разрабатывать новые методы, модели и алгоритмы для работы с большими данными, адаптируя их под конкретные задачи и требования (ПК-6);

аналитическая деятельность:

- Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика, интерпретировать результаты и формулировать выводы (ПК-7);
- Способен проводить испытания и разрабатывать рекомендации по внедрению новых методов и технологий для повышения эффективности работы с большими данными (ПК-8)

Проектная деятельность:

- Способен разрабатывать и согласовывать технические задания на создание методической и технологической инфраструктуры больших данных, обеспечивая соответствие проектных решений требованиям бизнеса и технологии (ПК-9).

5.2. Программа направлена на приобретение профессиональных компетенций, необходимых для выполнения трудовых функций:

Трудовая функция с кодом[1]	Профессиональные компетенции, обеспечивающие выполнение трудовой функции	
	индекс	содержание компетенции
Выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных (A/01.6)	ПК-1	Способен выявлять, формировать и согласовывать требования к результатам аналитических работ с применением технологий больших данных
Планирование и организация аналитических работ с использованием технологий больших данных (A/02.6)	ПК-3	Способен планировать и организовывать аналитические работы с использованием технологий больших данных, включая разработку расписания и распределение задач
Подготовка данных для проведения аналитических работ по исследованию больших данных (A/03.6)	ОПК-2	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	ПК-2	Способен подготовить данные для проведения аналитических работ по исследованию больших данных, включая их получение и очистку
Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика (A/04.6)	УК-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ях)
	ОПК-1	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности
	ОПК-3	Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности

[1] Указываются все трудовые функции из раздела 4.2.

	ПК-7	Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика, интерпретировать результаты и формулировать выводы
Анализ потребности заинтересованных лиц и/или подразделений организации в исследовании больших данных (B/01.7)	УК-5	Способен воспринимать межкультурное разнообразие общества в социально-историческом, этическом и философском контекстах
	ПК-4	Способен анализировать потребности заинтересованных лиц и подразделений в исследовании больших данных и формировать предложения по их удовлетворению
Управление получением, хранением, передачей, обработкой больших данных (B/05.7)	ПК-5	Способен управлять процессами получения, хранения, передачи и обработки больших данных, включая мониторинг и поддержку инфраструктуры
Совершенствование и разработка новых методов, моделей, алгоритмов и инструментальных средств (D/01.8)	ОПК-8	Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем
	ПК-6	Способен совершенствовать и разрабатывать новые методы, модели и алгоритмы для работы с большими данными, адаптируя их под конкретные задачи и требования
Проведение испытаний и разработка рекомендаций по внедрению новых методов, моделей и технологий (D/02.8)	ОПК-4	Способен участвовать в разработке технической документации, связанной с профессиональной деятельностью с использованием стандартов, норм и правил
	ПК-8	Способен проводить испытания и разрабатывать рекомендации по внедрению новых методов и технологий для повышения эффективности работы с большими данными

6. УЧЕБНЫЙ ПЛАН

Специалист по анализу данных Data Scientist

Трудоёмкость 560 академических часов. Срок освоения 14 недель. Форма обучения очно-заочная, с применением электронного обучения и дистанционных образовательных технологий.

Цель: получение компетенций для анализа данных, построения, проверки и практического применения моделей машинного обучения, обработки больших данных и представления результатов.

К освоению допускаются лица, имеющие или получающие СПО и (или) высшее образование и владеющие базовым синтаксисом Python.

Л - лекции; П - практика с преподавателем; А - контактная итоговая аттестация; С - самостоятельная работа. Все значения в академических часах.

Тема	Л	П	А	С	Всего
1.1 Настройка среды и программирование на Python	10	20	0	10	40
1.2 Исследование данных и описательная статистика	5	20	0	15	40
1.3 Продакшен пет-проекта	5	20	0	15	40
2.1 Теория вероятностей, статистика и А/Б тестирование	8	15	0	7	30
2.2 Линейная алгебра и анализ	8	15	0	7	30
2.3 Машинное обучение с учителем	7	15	0	8	30
2.4 Кластеризация, временные ряды	7	15	0	8	30

6. УЧЕБНЫЙ ПЛАН

Тема	Л	П	А	С	Всего
3.1 Введение в нейронные сети и компьютерное зрение	10	20	0	10	40
3.2 Обработка и анализ изображений: детекция, сегментация	10	20	0	10	40
3.3 Анализ текстов с помощью нейронных сетей	10	20	0	10	40
3.4 Рекомендательные системы и мультимодальные задачи	4	26	0	10	40
4.1 SQL и PySpark	8	22	0	10	40
4.2 Визуализация и автоматизация работы с данными	6	24	0	10	40
5 Подготовка выпускного проекта	0	0	0	40	40
6 Итоговая аттестация	0	0	4	36	40

Сводный баланс

Раздел	Л	П	А	С	Всего
1 Среды разработки и основы Python	20	60	0	40	120
2 Статистика и классическое машинное обучение	30	60	0	30	120
3 Нейросетевое моделирование	34	86	0	40	160
4 Обработка больших данных	14	46	0	20	80
Подготовка проекта и итоговая аттестация	0	0	4	76	80
Всего	98	252	4	206	560

Контактная работа: 354 часа. Самостоятельная работа: 206 часов. Итоговая аттестация включена в 560 часов. Контроль по модулям 1-4 - практическая работа; текущий контроль по темам - выполнение заданий и объяснение решения. Компетенции соответствуют разделу 5 и рабочим программам раздела 8.

7. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Период освоения: 14 недель, 70 учебных дней. Каждая неделя включает 40 академических часов; каждый учебный день - 8 академических часов общей нагрузки. Даты и время занятий закрепляются в расписании потока.

Нед.	Дни	Содержание	Л	П	А	С	Всего
1	1-5	1.1	10	20	0	10	40
2	6-10	1.2	5	20	0	15	40
3	11-15	1.3	5	20	0	15	40
4	16-20	2.1 полностью; 2.2 первые 10 ч	11	20	0	9	40
5	21-25	2.2 оставшиеся 20 ч; 2.3 первые 20 ч	10	20	0	10	40
6	26-30	2.3 оставшиеся 10 ч; 2.4 полностью	9	20	0	11	40
7	31-35	3.1	10	20	0	10	40
8	36-40	3.2	10	20	0	10	40
9	41-45	3.3	10	20	0	10	40
10	46-50	3.4	4	26	0	10	40
11	51-55	4.1	8	22	0	10	40
12	56-60	4.2	6	24	0	10	40
13	61-65	Подготовка выпускного проекта	0	0	0	40	40
14	66-70	Итоговая аттестация	0	0	4	36	40

Л - лекции; П - практика; А - контактная аттестация; С - самостоятельная работа. Занятия проводятся дистанционно.

Переходящие темы: тема 2.2 на неделе 4 - 3/5/2 часа Л/П/С, на неделе 5 - 5/10/5; тема 2.3 на неделе 5 - 5/10/5, на неделе 6 - 2/5/3. Это обеспечивает по 30 часов на каждую тему второго модуля.

8.РАБОЧАЯ ПРОГРАММА

8.1. Трудоемкость освоения: 560 часов

8.2. Задачи:

1. Освоение основ Python и среды разработки: слушатели должны научиться настраивать рабочее окружение для программирования и освоить базовые техники разработки на Python, что позволит эффективно работать с данными и разрабатывать свои первые проекты (pet-проекты).
2. Применение методов описательной статистики для анализа данных: обучение основам сбора, обработки и визуализации данных с помощью Python, а также базовым статистическим методам для описания и интерпретации полученных данных.
3. Изучение основ теории вероятностей и статистики для анализа данных и проведения А/Б тестирования: слушатели приобретут навыки работы с вероятностными моделями и статистическими методами, что даст возможность грамотно оценивать гипотезы и проводить эксперименты, например, А/Б тесты.
4. Применение линейной алгебры и анализа для решения задач машинного обучения: участники научатся использовать математические основы для построения моделей машинного обучения с учителем, включая регрессию, классификацию и работу с временными рядами.
5. Изучение методов кластеризации и временных рядов: слушатели освоят методы безучительного обучения и смогут анализировать данные, распределённые по временным интервалам.
6. Изучение нейросетевого моделирования для решения сложных задач: участники познакомятся с основами построения нейронных сетей, будут работать с рекомендательными системами и мультимодальными задачами.
7. Применение SQL и PySpark для работы с большими данными: слушатели приобретут навыки работы с большими массивами данных с использованием SQL и PySpark, что даст возможность интегрировать и анализировать данные в средах с большими объёмами информации.
8. Интеграция полученных знаний через продакшен собственного проекта: участники курса смогут применить полученные знания на практике, разработав и завершив свой pet-проект, который продемонстрирует их компетенции в сфере анализа данных и машинного обучения.

8.3. Учебная (рабочая) программа:

№ п/п, название темы	Виды учебных занятий, учебных работ	Содержание	Планируемые результаты обучения (Знать/Уметь/Владеть)
Раздел 1. Среды разработки и основы Python			
1.1. Настройка	Лекции (10 ч.)	Установка и настройка сред работы с языком Python.	**Знает:**

среды и программирование на Python		Работа с интерактивной средой разработки Jupyter Notebook.	Основные среды разработки программного обеспечения (IDE) и их особенности. Основные концепции языка программирования Python (синтаксис, переменные, типы данных, операторы, условия, циклы). Принципы обработки данных: получение, очистка, предварительный анализ данных. Современные информационные технологии, используемые в профессиональной деятельности. Основные правила информационной безопасности при работе с данными. **Умеет:** Настраивать и работать в различных интегрированных средах разработки (PyCharm, Visual Studio Code, Jupyter Notebook). Писать простые программы на языке Python для решения типовых задач. Обрабатывать и очищать данные с использованием
	Практические занятия (20 ч.)	Настройка и запуск сред разработки. Решение задач на типы данных, функции и классы языка программирования Python. Введение в язык bash.	
	Самостоятельная работа (10 ч.)	Интеграция платформ разработки с системами хранения и версионирования кода.	
1.2. Исследование данных и описательная статистика	Лекции (5 ч.)	1. Основные операции при обработке табличных типов данных (группировки, агрегации, фильтрация) 2. Визуализация данных и построение диаграмм (гистограммы, диаграммы рассеяния, столбчатые диаграммы)	
	Практические занятия (20 ч.)	Применение библиотек pandas, matplotlib и seaborn для обработки, фильтрации и визуализации данных.	
	Самостоятельная работа (15 ч.)	Применение библиотеки plotly для продвинутой визуализации данных.	
1.3. Продакшен пет-проекта	Лекции (5 ч.)	1. Основные принципы разработки чат-ботов на платформе Telegram 2. Разработка веб-сервисов с использованием streamlit.	
	Практические занятия (20 ч.)	Реализация telegram-бота для транслитерации текстов. Разработка веб-сервиса визуализации данных. Развертывание проекта на внешних серверах.	
	Самостоятельная работа (15 ч.)	Контейнеризация готового веб-приложения с использованием Docker.	

			библиотек Python (pandas, numpy). Применять базовые методы анализа данных. Использовать информационные технологии и программные средства для решения профессиональных задач. **Владеет:** Практическими навыками работы в современных средах разработки. Навыками работы с библиотеками для обработки данных в Python. Навыками решения стандартных задач профессиональной деятельности с использованием информационно-коммуникационных технологий. Навыками получения и предварительной очистки данных для аналитических работ.
Раздел 2. Статистика и классическое машинное обучение			
2.1. Теория вероятностей, статистика и А/Б	Лекции (8 ч.)	1. Введение в теорию вероятностей: основные понятия 2. Основные статистические распределения 3. Проверка статистических гипотез: выбор теста, объем выборки, статистические критерии 4. А/Б тестирование	**Знает:** Основные понятия и методы статистического анализа данных. Принципы построения и применения классических алгоритмов
	Практическая	Решение задач на базовые методы теории вероятностей и статистики.	

тестирование	работа (15 ч.)	Проверка гипотез с помощью параметрических и непараметрических тестов. Визуализация результатов проверки гипотез.	машинного обучения (линейная регрессия, логистическая регрессия, методы ближайших соседей, деревья решений и т.д.). Методы математического анализа и моделирования для решения задач обработки данных. Основные этапы проведения аналитических исследований с применением технологий больших данных. Способы выявления и формулирования требований к аналитическим работам с учетом целей заказчика. **Умеет:** Применять статистические методы анализа данных для выявления закономерностей. Строить и обучать классические модели машинного обучения для решения задач классификации, регрессии и кластеризации. Планировать и организовывать аналитические исследования, включая определение целей, выбор методов анализа и моделирования.
	Самостоятельная работа (7 ч.)	Разработка веб-приложения для визуализации различных распределений.	
2.2. Линейная алгебра и анализ	Лекции (8 ч.)	1. Введение в линейную алгебру: понятие вектора и операции с векторами 2. Работа с матрицами и матричные разложения 3. Введение в математический анализ: понятие производных и производные элементарных функций 4. Основы градиентного спуска и оптимизации моделей. 5. Линейная и логистическая регрессия.	
	Практическая работа (15 ч.)	Решение задач на базовые векторные и матричные операции. Применение матричных разложений при работе с изображениями. Реализация градиентного спуска для одной и двух переменных.	
	Самостоятельная работа (7 ч.)	Реализация классов градиентного спуска для линейной и логистической регрессии с использованием собственных компонентов.	
2.3. Машинное обучение с учителем	Лекции (7 ч.)	1. Базовые методы машинного обучения: метод ближайшего соседа, машина опорных векторов, деревья решений 2. Ансамблирование моделей: бэггинг, стекинг, голосование, бустинг 3. Метрики качества моделей 4. Отбор признаков	
	Практическая работа (15 ч.)	Применение базовых моделей для решения задачи классификации. Применение ансамблевых методов для решения задач классификации и регрессии. Сравнение моделей по различным наборам метрик.	
	Самостоятельная работа (8 ч.)	Решение задачи регрессии цен на недвижимость.	
2.4.. Класт	Лекции (7 ч.)	1. Методы кластеризации и метрики в задаче кластеризации.	

ериза ция, време нные ряды		2. Введение в обработку текста и кластеризация текстовых данных. 3. Введение во анализ временных рядов и применение базовых статистических методов.	Анализировать потребности заинтересованных лиц в исследованиях больших данных и формировать предложения по их удовлетворению. Интерпретировать результаты аналитических исследований и формулировать выводы на основе полученных данных. **Владеет:** Навыками работы с методами статистического анализа и моделирования для решения профессиональных задач. Навыками работы с инструментами и библиотеками для машинного обучения (например, scikit-learn, pandas). Навыками выбора и применения оптимальных методов анализа данных в зависимости от требований заказчика. Навыками структурированного представления результатов анализа и их интерпретации для различных заинтересованных сторон. Навыками согласования требований к
	Практическая работа (15 ч.)	Кластеризация состояний мобильного устройства на основе показаний датчиков. Кластеризация новостей для задачи тематического моделирования. Анализ продаж товаров с использованием методов прогнозирования временных рядов.	
	Самостоятельная работа (8 ч.)	Решение задачи кредитного скоринга.	

			аналитическим работам с применением технологий больших данных.
Раздел 3. Нейросетевое моделирование			
3.1. Введение в нейронные сети и компьютерное зрение.	Лекции (10 ч.)	1. Введение в нейросетевое моделирование. Полносвязные слои, регуляризация, pytorch. 2. Оптимизация нейронных сетей. 3. Сверточные слои для анализа изображений. 4. Большие сверточные архитектуры.	**Знает:** Основные принципы работы нейронных сетей и их архитектур (перцептрон, рекуррентные, сверточные и другие типы сетей). Методы обучения нейронных сетей (градиентный спуск, обратное распространение ошибки, стохастический градиентный спуск и т.д.). Способы применения нейросетевых моделей для решения задач прогнозирования, классификации и регрессии. Основные методы работы с большими данными в контексте нейросетевого моделирования. Математические модели и алгоритмы, используемые для проектирования нейронных сетей. **Умеет:** Планировать и
	Практическая работа (20 ч.)	Решение задачи регрессии с помощью полносвязных нейронных сетей. Анализ различных алгоритмов оптимизации (импульсные и адаптивные алгоритмы градиентного спуска). Решение задачи классификации изображений с помощью сверточных нейронных сетей. Дообучение больших сверточных архитектур для применения моделей на собственных данных.	
	Самостоятельная работа (10 ч.)	Реализация и развертывание веб-приложения для классификации изображений с помощью простых и дообученных нейронных сетей.	
3.2. Обработка и анализ изображений: детекция, сегментация	Лекции (10 ч.)	1. Постановка задачи локализации объектов на изображении. 2. Задача детекции объектов на изображении и одноэтапные модели детекции. 3. Задача сегментации и двухэтапные модели компьютерного зрения.	
	Практическая работа (20 ч.)	Реализация собственной модели для локализации объектов на изображении. Применение одноэтапной модели детекции и разметка данных. Применение двухэтапных моделей детекции для решения задач детекции и сегментации.	
	Самостоятельная работа (10 ч.)	Реализация и развертывание веб-приложения для решения задач детекции и сегментации объектов на изображении.	
3.3. Анализ текстов с помощью нейро	Лекции (10 ч.)	1. Векторное представление текста и введение в рекуррентные нейронные сети. 2. Применение механизма внимания для повышения эффективности рекуррентных моделей. 3. Введение в архитектуру трансформер и применение модели BERT. 4. Декодер и генерация текста.	

нных сетей	Практическая работа (20 ч.)	Применение рекуррентных моделей для классификации текста. Сравнение моделей с механизмом внимания и простых рекуррентных моделей для задачи обработки текста.	организовывать аналитические работы с применением технологий больших данных в процессе создания нейросетевых моделей. Разрабатывать и обучать нейросетевые модели для решения специфических задач с использованием технологий глубокого обучения (deep learning). Совершенствовать существующие нейросетевые модели и алгоритмы, адаптируя их для конкретных задач и требований. Применять математические методы и модели для проектирования эффективных нейросетевых решений. Проводить анализ производительности нейросетевых моделей и оптимизировать их параметры для повышения точности и скорости. **Владеет:** Навыками работы с современными фреймворками для создания и обучения нейронных сетей (TensorFlow, PyTorch,
	Самостоятельная работа (10 ч.)	Реализация и развертывание веб-приложения для для решения задачи оценки тональности текста и генерации текста.	
3.4. Рекомендательные системы и мультимодальные задачи	Лекции (4 ч.)	1. Классические метрические подходы в задаче рекомендаций. 2. Нейросетевые подходы в задаче рекомендаций и метрики качества рекомендаций.	
	Практическая работа (26 ч.)	Решение задачи рекомендации фильмов с помощью классических подходов. Решение задачи рекомендации фильмов с помощью нейросетевых подходов и реализация метрик рекомендаций.	
	Самостоятельная работа (10 ч.)	Реализация и развертывание веб-приложения поисково-рекомендательной системы на различных текстовых данных.	

			Keras и др.). Навыками планирования и распределения задач при реализации проектов, связанных с нейросетевым моделированием. Навыками оптимизации архитектуры нейросетей, включая настройку гиперпараметров и выбор подходящих алгоритмов обучения. Навыками анализа результатов работы нейросетей и интерпретации их предсказаний.
Раздел 4. Обработка больших данных			
4.1. SQL & PySpark	Лекции (8 ч.)	1. Введение в язык запросов SQL. 2. Работа с PostgresSQL. 3. PySpark и распределенная обработка данных.	**Знает:** Основные принципы работы с большими данными, включая архитектуру систем обработки данных (Hadoop, Spark). Основные функции и возможности языка запросов SQL для работы с реляционными базами данных. Принципы работы с распределенными вычислениями и обработкой данных в PySpark. Методы получения, хранения, передачи и обработки данных в
	Практическая работа (22 ч.)	Использование языка SQL для фильтрации, группировки, агрегации данных. Установка PostgresSQL на облачных сервера, удаленное подключение и экспорт данных. Установка и подключение PySpark, анализ данных.	
	Самостоятельная работа (10 ч.)	Решение задачи регрессии цен на недвижимость с помощью PySpark и MLLib.	
4.2. Визуализация и автоматизация работы с данными	Лекции (6 ч.)	1. Установка, настройка и подключение к источнику данных с помощью Apache Superset. 2. Построение витрин данных с помощью Apache Superset. 3. Настройка автоматизированных процессов сбора данных с помощью Apache Airflow.	
	Практическая работа (24 ч.)	Развертывание Apache Superset в облачной среде и разработка витрины данных. Развертывание и настройка автоматизированных процессов сбора и обработки данных.	

	Самостоятельная работа (10 ч.)	Построение инфраструктуры сбора, обработки и визуализации данных в облачной среде.	системах больших данных. **Умеет:** Управлять процессами получения, хранения и передачи данных, используя SQL и PySpark. Применять SQL для извлечения и трансформации данных в реляционных базах данных. Использовать PySpark для обработки и анализа больших объемов данных в распределенной среде. Разрабатывать техническую документацию, соответствующую стандартам и нормам, в процессе реализации проектов с большими данными. Анализировать результаты обработки больших данных и предлагать рекомендации по улучшению процессов и внедрению новых технологий. **Владеет:** Навыками работы с системами управления базами данных (DBMS), использующими SQL для обработки больших объемов данных.

			Навыками распределенной обработки данных с использованием PySpark для оптимизации производительности. Навыками мониторинга и поддержки инфраструктуры больших данных, включая процессы хранения и передачи данных. Навыками тестирования и оценки новых методов и технологий для повышения эффективности обработки данных. Навыками документирования процессов и разработки рекомендаций по внедрению новых технологий в рамках проектов с большими данными.

8.4. Текущая аттестация.

Текущий контроль успеваемости и качества подготовки осуществляется в пределах времени, отведенного на учебные занятия, и выполняет одновременно обучающую функцию. Текущий контроль успеваемости проводится в процессе изучения каждого раздела (темы, подтемы) внутри модуля данной Программы и проводится как в форме устного опроса преподавателем, так и практического написания программного кода.

Промежуточная аттестация проводится дистанционно в форме выполнения практического задания и объяснения решения преподавателю. Итоговая аттестация проводится в форме дипломного итогового проекта с представлением результата и ответами на вопросы в соответствии с разделом 10.

По окончании каждого модуля рабочей программы проводится промежуточная аттестация в виде практической работы, по теме каждого модуля данной Программы.

Практическая работа

Форма проведения	С применением ДОТ
Виды оценочных материалов	Практические задания в модуле

Критерии оценивания	1. 1-3 балла – Задание не выполнено или результатами работы невозможно воспользоваться на практике 2. 4-7 баллов – Есть большие неточности в выполнении задания 3. 8-10 баллов – Задача решена полностью, результат можно использовать на практике
Оценка	Для того чтобы успешно сдать работу, достаточно набрать 7 баллов.

8.5. Оценочные средства промежуточной аттестации обучающихся по освоению рабочей программы

Примерная тематика контрольных вопросов, выявляющих теоретическую подготовку обучающегося:

- В чем различие между независимыми и условными событиями?
- Какое распределение описывает количество успехов в фиксированном числе независимых испытаний?
- В чем ключевое отличие нормального распределения от экспоненциального?
- Каковы основные различия между задачами классификации и регрессии?
- Что такое гиперпараметры, и как они влияют на обучение модели?
- Как работает алгоритм K-means для кластеризации данных?
- В чем преимущество алгоритма DBSCAN по сравнению с K-means?
- Что такое функция активации, и какую роль она играет в нейронных сетях?
- Как осуществляется обратное распространение ошибки в нейронных сетях?
- В чем основная цель свертки в сверточных нейронных сетях (CNN)?
- Что такое фильтры (kernels) в сверточных слоях и как они используются?
- Какова основная идея рекуррентных нейронных сетей (RNN) и как они работают с последовательностями данных?
- Чем архитектура LSTM отличается от стандартного рекуррентного слоя?
- Как механизм внимания позволяет трансформеру обрабатывать длинные последовательности данных?
- В чем различие между «self-attention» и «multi-head attention» в архитектуре трансформера?
- Как с помощью SQL выбрать только уникальные записи из таблицы?
- Чем различаются команды `INNER JOIN` и `LEFT JOIN` в SQL?

Примеры задач, выявляющих уровень сформированности компетенций:

- Проверьте гипотезу о том, что средняя зарплата сотрудников компании A больше, чем средняя зарплата сотрудников компании B, используя тест Стьюдента на уровне значимости 5%.
- Постройте модель линейной регрессии для предсказания стоимости жилья на основе данных о количестве комнат, площади, расположении и других характеристиках дома.
- Используя алгоритм K-means, разделите клиентов интернет-магазина на группы по схожести их покупательских предпочтений для последующего создания персонализированных предложений.
- Обучите сверточную нейронную сеть для классификации изображений кошек и собак из набора данных, содержащего тысячи размеченных фотографий двух классов.
- Напишите SQL-запрос, который выводит количество продаж каждого продукта за последний год, а также их общую стоимость, отсортированную по убыванию.

8.6. Литература

Основная[9]*:*

1. Кострикин А.И. Линейная алгебра и геометрия. — М.: Наука, 2020. — 512 с. — ISBN 978-5-02-014533-2.
2. Кочетков Е.М., Смирнов С.П. Линейная алгебра: Учебное пособие. — М.: Физматлит, 2019. — 416 с. — ISBN 978-5-9221-1771-8.
3. Ширяев А.Н. Вероятность и математическая статистика. — М.: МЦНМО, 2019. — 592 с. — ISBN 978-5-4439-1696-5.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 2020. — 480 с. — ISBN 978-5-06-005857-8.
5. Гудфеллоу И., Бенджио Й., Курвилль А. Глубокое обучение. — М.: Издательство «Вильямс», 2018. — 800 с. — ISBN 978-5-8459-2006-9.
6. Мюллер А., Гвидо С. Введение в машинное обучение с Python: руководство для специалистов по обработке данных. — М.: ДМК Пресс, 2017. — 416 с. — ISBN 978-5-97060-378-6.
7. Бен-Форта Б. Изучаем SQL. — СПб.: Питер, 2021. — 480 с. — ISBN 978-5-4461-1615-1.
8. Синха Б. Архитектура облачных вычислений: Теория и практика. — М.: ДМК Пресс, 2020. — 352 с. — ISBN 978-5-97060-855-2.

Дополнительная[10]*:*

1. **Гельфанд И.М., Фомин С.В.** Лекции по линейной алгебре. — М.: Физматлит, 2018. — 320 с. — ISBN 978-5-9221-1725-1.
2. **Елизаров А.А.** Математическая статистика. — М.: Юрайт, 2021. — 416 с. — ISBN 978-5-534-04333-2.

Интернет-ресурсы:

1. Khan Academy (Алгебра) ресурс: https://www.khanacademy.org/math/algebra;
2. Statistical Analysis Resources by NIST (Математическая статистика) ресурс: https://www.itl.nist.gov/div898/handbook/;
3. Towards Data Science (Машинное обучение) ресурс: https://towardsdatascience.com/;
4. Linear Algebra by MIT OpenCourseWare (Алгебра) ресурс: https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/.

9

10

9. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

9.1. Организация образовательной деятельности по освоению программы: При реализации программы используются дистанционные образовательные технологии.
Учебный процесс происходит с использованием разнообразных методов организации и осуществления учебно-познавательной деятельности (словесные, наглядные и практические методы передачи информации, проблемные лекции и др.); стимулирования и мотивации учебно-познавательной деятельности (дискуссии и др.); контроля и самоконтроля. Используются активные формы проведения занятий: ситуационный анализ, ролевые игры, эвристические технологии, тренинги.

9.2. Продолжительность занятий и недельная нагрузка

Нагрузка составляет 40 академических часов в неделю, включая контактную и самостоятельную работу. Академический час - 45 минут. При пятидневной учебной неделе общая нагрузка составляет 8 академических часов в учебный день. Время занятий, самостоятельной работы и перерывов устанавливается расписанием потока. При полном учебном дне предусматриваются два перерыва не менее 10 минут и перерыв на обед не менее 90 минут. Перерывы в учебные часы не включаются.

9.3. Материально-техническая база реализации программы

Материально-техническая база соответствует действующим противопожарным правилам и нормам и обеспечивает проведение всех видов занятий и практической работы, предусмотренных учебным планом.

Компьютер слушателя: процессор 4 ядра, 2,7 ГГц; оперативная память 8 Гб; диск 1 Тб; монитор 21,5 дюйма, разрешение 1920×1080; ОС Ubuntu; архитектура x64. Используются SSD и процессоры с оценкой от 2500 и выше. Для дистанционного взаимодействия необходимы доступ в Интернет и средства аудио- и видеосвязи.

9.4. Квалификация педагогических работников

Квалификация педагогических работников соответствует применимым квалификационным требованиям, квалификационным характеристикам должностей работников образования и профессиональным стандартам при их наличии. Преподаватели владеют содержанием разделов, методами практического обучения и технологиями дистанционного взаимодействия.

10. ПРОГРАММА ИТОГОВОЙ АТТЕСТАЦИИ

К итоговой аттестации допускаются обучающиеся, освоившие учебные модули в объёме, предусмотренном программой. Итоговая аттестация проводится по завершении всей программы независимо от применения дистанционных технологий.

10.1. Форма и объём итоговой аттестации

Форма аттестации - дипломный итоговый проект (выпускной проект) с представлением результата и ответами на вопросы. На тринадцатой неделе выполняется подготовка проекта: 40 часов самостоятельной работы. На четырнадцатой неделе проходит итоговая аттестация: 40 часов, из них 36 часов самостоятельной работы над итоговыми материалами и 4 часа контактной работы с преподавателем. Все 80 часов входят в общие 560 часов программы.

Этапы работы над проектом

1. Анализ предметной области и сбор данных.
2. Очистка и валидация собранных данных.
3. Анализ моделей и выбор решения задачи.
4. Разработка архитектуры приложения с моделью машинного обучения.
5. Реализация веб-приложения или чат-бота и подготовка презентации.

Результат объединяет материалы модулей программы и представляет собой работающий прототип MVP для использования в сети Интернет.

Требования к результату

Для сбора данных применяются подходящие библиотеки, например Beautiful Soup, Scrapy или Selenium. При необходимости размечаются графические данные. Модель реализуется с помощью scikit-learn, PyTorch или аналогов. Для MVP применяются Streamlit, Flask, Gradio, aiogram или аналогичные средства. Исходный код и версии проекта хранятся в Git.

Оценивание

Оценивание проводится по критериям раздела 11.3 по 100-балльной шкале. Итог «зачтено» соответствует 70-100 баллам (оценки 3, 4 или 5); «не зачтено» - 69 баллам и менее (оценка 2). Решение и замечания фиксируются в результатах аттестации. Порядок повторной аттестации определяется локальными актами организации.

Обучающимся, успешно освоившим программу и прошедшим итоговую аттестацию, выдаётся диплом о профессиональной переподготовке в соответствии с пунктом 3.6 настоящей программы.

11. ОЦЕНОЧНЫЕ СРЕДСТВА ИТОГОВОЙ АТТЕСТАЦИИ

11.1. Примерная тематика контрольных вопросов, выявляющих теоретическую подготовку обучающегося:

1. Статистические распределения и их применения в проверке гипотез.
2. Линейные методы в машинном обучении.
3. Деревья решений и ансамбли моделей.
4. Методы нормализации данных и случаи их применения.
5. Методы кодирования категориальных данных.
6. Простейшие нейронные сети.
7. Сверточные сети, нормализация, аугментация данных.
8. Рекуррентные сети и векторные представления текста.
9. Трансформеры в обработке текста.
10. Рекомендательные системы и семантический поиск.
11. Работа с реляционными базами данных. Обработка больших данных.
12. Построение интерфейсов прикладного программирования и их развертывание в облачной среде.
13. Управление данными с помощью инструментов оркестрации.
14. Построение витрин данных.
15. Контейнеризация приложений.

11.2. Примеры заданий, выявляющих практическую подготовку обучающегося

1. Проведение статистических тестов для связных и несвязных выборок.
2. Оценка необходимого объема выборок для статистического исследования.
3. Оценка качества моделей с помощью соответствующих метрик классификации и регрессии.
4. Применение перекрестной проверки в процессе обучения и сравнения моделей.
5. Оптимизация гиперпараметров моделей.
6. Применение сверточных нейронных сетей в классификации картинок.
7. Применение трансформеров и рекуррентных моделей в задаче классификации текста.
8. Применение языка структурных запросов в работе с реляционными базами данных.
9. Разработка и развертывание веб-сервисов для решения задач машинного обучения.

11.3. Критерии оценки при 100-балльной системе

Характеристика ответа	Баллы	Оценка

Дан полный, развернутый ответ на поставленный вопрос, показана совокупность осознанных знаний об объекте, проявляющаяся в свободном оперировании понятиями, умении выделить существенные и несущественные его признаки, причинно-следственные связи. Знание об объекте демонстрируется на фоне понимания его в системе данной науки и междисциплинарных связей. Ответ формулируется в терминах науки, изложен литературным языком, логичен, доказателен, демонстрирует авторскую позицию обучающегося. Практические (и/или лабораторные) работы выполнены в полном объеме, теоретическое содержание курса освоено полностью, необходимые практические навыки работы в рамках учебных заданий сформированы, все предусмотренные программой учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному	90-100	5
Дан полный, развернутый ответ на поставленный вопрос, показана совокупность осознанных знаний об объекте, доказательно раскрыты основные положения темы; в ответе прослеживается четкая структура, логическая последовательность, отражающая сущность раскрываемых понятий, теорий, явлений. Знание об объекте демонстрируется на фоне понимания его в системе данной науки и междисциплинарных связей. Ответ изложен литературным языком в терминах науки. Могут быть допущены недочеты в определении понятий, исправленные обучающимся самостоятельно в процессе ответа или с помощью преподавателя. Практические (и/или лабораторные) работы выполнены в полном объеме, теоретическое содержание курса освоено полностью, необходимые практические навыки работы в рамках учебных заданий в основном сформированы, все предусмотренные программой обучения учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному	80-89	4
Дан недостаточно полный и недостаточно развернутый ответ. Логика и последовательность изложения имеют нарушения. Допущены ошибки в раскрытии понятий, употреблении терминов. Обучающийся не способен самостоятельно выделить существенные и несущественные признаки и причинно-следственные связи. Обучающийся может конкретизировать обобщенные знания, доказав на примерах их основные положения только с помощью преподавателя. Речевое оформление требует поправок, коррекции. Практические (и/или лабораторные) работы выполнены, теоретическое содержание курса освоено частично, необходимые практические навыки работы в рамках учебных заданий в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки	70-79	3

Дан неполный ответ, представляющий собой разрозненные знания по теме вопроса с существенными ошибками в определениях. Присутствуют фрагментарность, нелогичность изложения. Обучающийся не осознает связь данного понятия, теории, явления с другими объектами дисциплины. Отсутствуют выводы, конкретизация и доказательность изложения. Речь неграмотная. Дополнительные и уточняющие вопросы преподавателя не приводят к коррекции ответа обучающегося не только на поставленный вопрос, но и на другие вопросы дисциплины. Практические (и/или лабораторные) работы выполнены частично, теоретическое содержание курса освоено частично, необходимые практические навыки работы в рамках учебных заданий не сформированы, большинство предусмотренных программой обучения учебных заданий не выполнено либо качество их выполнения оценено числом баллов близким к минимальному. При дополнительной самостоятельной работе над материалом курса, при консультировании преподавателя, возможно повышение качества выполнения учебных заданий	69 и менее	2

12. ИНЫЕ КОМПОНЕНТЫ ПРОГРАММЫ

12.1. Кадровое обеспечение образовательного процесса

Реализацию программы обеспечивают педагогические работники, соответствующие требованиям пункта 9.4. Состав преподавателей, распределение модулей и расписание определяются организацией для конкретного учебного потока.

Преподаватели проводят занятия, обеспечивают обратную связь по практическим заданиям, осуществляют текущий контроль и аттестацию. Учебные материалы и задания предоставляются слушателям через электронную образовательную среду организации.

Сведения о преподавателях и назначении на учебный поток фиксируются в организационных документах. Изменение состава преподавателей при сохранении установленных квалификационных требований не изменяет содержание и трудоёмкость программы.