Документ подписан квалифицированной электронной подписью
Сертификат: 025B3AE00099B2A2B548158CFF985138BF
Владелец: "АНО ВО ""РОССИЙСКИЙ НОВЫЙ УНИВЕРСИТЕТ"", ДИ
Действителен: с 07.03.2025 по 07.06.2026

Автономная некоммерческая организация
высшего образования
«Российский новый университет»
(АНО ВО «Российский новый университет»)

УТВЕРЖДАЮ
Первый проректор
_____ Е.В. Лобанова
«18» августа 2025 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА - ПРОГРАММА ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ

«Архитектор данных с фокусом на машинное обучение» (при работе с большими данными)

Объем программы: 864 часа

Москва
2025

1. Общая характеристика программы

1.1. Дополнительная профессиональная программа – профессиональной переподготовки направлена на получение компетенции, необходимой для выполнения нового вида профессиональной деятельности по организации деятельности обучающихся по освоению знаний, формированию и развитию умений и компетенций, позволяющих осуществлять профессиональную деятельность, обеспечение достижения ими нормативно установленных результатов в сфере информационной безопасности (ИБ) направленной на использование технологий искусственного интеллекту при разработке алгоритмов и программно-технических средств, решающих задачи информационной безопасности в различных предметных областях

1.2. Нормативно-правовые акты, регламентирующие разработку дополнительной образовательной программы:

– Федеральный закон от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации»;

– Порядок организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденный приказом Минобрнауки России от 24 марта 2025 г. № 266;

– Устав Автономной некоммерческой организации высшего образования «Российский новый университет»;

– Локально - нормативные акты, регламентирующие образовательную деятельность по дополнительным образовательным программам.

1.3. Программа профессиональной переподготовки разработана с учетом требований:

- Приказ Министерства труда и социальной защиты Российской Федерации от 06.07.2020 № 405н "Об утверждении профессионального стандарта "Специалист по большим данным" (Зарегистрирован 05.08.2020 № 59174);
- Федерального государственного образовательного стандарта высшего образования - по направлению подготовки 09.04.03 Прикладная информатика, утвержденного приказом Министерства образования и науки Российской Федерации от 19.09.2017 N 916.

1.4. Категория обучающихся: лица, имеющие или получающие высшее.

Срок освоения программы: 864 часа (39 недели – 9 месяцев).

Режим обучения – до 5 дней в неделю, до 5 часов в день.

Форма обучения – заочная с применением дистанционных образовательных технологий.

Формы аттестации обучающихся: промежуточная и итоговая аттестация в форме зачетов и экзаменов.

Документ о квалификации: лицам, успешно освоившим программу профессиональной переподготовки и прошедшим итоговую аттестацию, выдается диплом о профессиональной переподготовке образца, установленного АНО ВО «Российский новый университет».

1.5. Цель программы: совершенствование компетенций слушателей, необходимых для их профессиональной деятельности в сфере машинного обучения (ИБ) направленной на использование технологий искусственного интеллекта при разработке алгоритмов и программно-технических средств, решающих задачи с большими данными в различных предметных областях.

Слушатель, освоивший программу профессиональной переподготовки, в соответствии с целью программы профессиональной переподготовки, на которые ориентирована программа, должен быть готов решать следующие **профессиональные задачи:**

1. Проведение анализа локальных и мировых потребностей в совершенствовании и разработке новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными.
2. Разработка методики и плана испытаний усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными.
3. Разработка рекомендаций по внедрению и использованию усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными.
4. Интеграция усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными с методической и технологической инфраструктурой больших данных.

2. Планируемые результаты обучения:

В результате освоения данной программы в соответствии с целью программы и задачами профессиональной деятельности, слушатель должен обладать всеми профессиональными компетенциями:

КОД	Наименование видов деятельности и профессиональных компетенций
ПК 1	Способен совершенствовать и разрабатывать новые методы, модели, алгоритмы, технологии и инструментальные средства работы с большими данными
ПК 2	Способен проводить анализ и оценку имеющегося научного задела и ресурсного потенциала научного подразделения
ПК 3	Способен контролировать ход выполнения работ: мониторинг, оценка объема и качества выполненных работ, корректировка отклонений от плана выполнения работ
ПК 4	Способен проводить испытания и разработку рекомендаций по внедрению и использованию усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными
ПК 5	Способен интегрировать усовершенствованные или разработанные новые методы, модели, алгоритмы, технологии и инструментальные средства работы с большими данными
ПК 6	Способен владеть управлением процессом испытаний
ПК 7	Способен применять современный опыт использования технологий больших данных

В результате освоения дополнительной профессиональной программы – программы профессиональной переподготовки слушатель должен **знать:**

1. Состояние и перспективы развития в Российской Федерации и в мире технологий больших данных и других информационных технологий для задач информационно-аналитической деятельности, поддержки принятия решений, а также создания инновационных продуктов и услуг;

2. Существующие и перспективные архитектуры больших данных, в том числе распределенные кластерные архитектуры;

3. Методы сравнительного анализа применительно к математическим методам, моделям, алгоритмам, технологиям и инструментальным средствам.

4. Современное состояние и прогноз развития технологий больших данных.

В результате освоения дополнительной профессиональной программы профессиональной переподготовки слушатель должен **уметь:**

1. Проводить аналитические исследования по тематике информационных технологий, технологий больших данных;

2. Разрабатывать конкурсную, проектную и рабочую документацию на проведение работ;

3. Проводить исследования и испытания заданных методов, моделей, алгоритмов, технологий и инструментальных средств в соответствии с методикой испытаний;

4. Подготавливать документацию для проведения патентных исследований, оформления прав на интеллектуальную собственность.

В результате освоения дополнительной профессиональной программы профессиональной переподготовки слушатель должен **иметь практический опыт:**

1. Проведения анализа локальных и мировых потребностей в совершенствовании и разработке новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными.;

2. Разработки методики и плана испытаний усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными.;

3. Разработки рекомендаций по внедрению и использованию усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными;

4. Интеграции усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными с методической и технологической инфраструктурой больших данных.

3. Формы аттестации

Формами аттестации слушателей по программе профессиональной являются: промежуточная и итоговая аттестация.

Промежуточная аттестация проводится в форме зачетов и экзаменов, предназначена для объективного подтверждения и оценивания достигнутых результатов обучения после завершения изучения дисциплины. Оценивание результатов формирования компетенций в рамках дисциплины у слушателей осуществляется по промежуточной аттестации.

Итоговая аттестация слушателей по программе профессиональной переподготовки включает итоговый экзамен, который проходит в форме решения практической задачи.

4. Документ об обучении (образовании)

Лицам, успешно освоившим дополнительную профессиональную программу и прошедшим итоговую аттестацию, выдаётся диплом о профессиональной переподготовки образца, установленного АНО ВО «РосНОУ».

При освоении дополнительной профессиональной программы параллельно с получением среднего профессионального образования и (или) высшего образования диплом о профессиональной переподготовке выдается одновременно с получением

соответствующего документа об образовании и о квалификации.

Лицам, не прошедшим итоговой аттестации или получившим на итоговой аттестации неудовлетворительные результаты, а также лица освоившим часть дополнительной профессиональной программы и (или) отчисленным из организации, выдается справка об обучении или о периоде обучения по образцу, установленного АНО ВО «РосНОУ».

5. Учебный план

УЧЕБНЫЙ ПЛАН
программы профессиональной переподготовки
«Архитектор данных с фокусом на машинное обучение» (при работе с большими данными)

№п/п	Наименование учебных курсов, дисциплин (модулей) практик	Всего час.	В том числе			Сам. раб	Форма контроля	Формируемые компетенции
			аудит. занят.	лекции	практич. зан.			
1	**Методы машинного обучения**	**324**	**162**	**8**	**10**	**162**		ПК 3, ПК 5, ПК 7
1.1	Фундаментальные алгоритмы машинного обучения	36	18	8	10	18	зачет	
1.2	Анатомия алгоритмов обучения	36	18	8	10	18	зачет	
1.3	Выбор алгоритма машинного обучения	36	18	8	10	18	зачет	
1.4	Нейронные сети и глубокое обучение	72	36	18	18	36	зачет	
1.5	Подготовка данных для алгоритмов машинного обучения	36	18	8	10	18	зачет	
1.6	Продвинутые методики машинного обучения	72	36	18	18	36	зачет	
1.7	Обучение без учителя	36	18	8	10	18	зачет	
2	**Семантическая обработка естественно-языковых текстов**	**252**	**126**	**52**	**74**	**126**		ПК 2, ПК 6
2.1	Методы обработки естественного языка	36	18	8	10	18	зачет	
2.2	Основные библиотеки для обработки естественного языка и их возможности	36	18	8	10	18	зачет	
2.3	Основные возможности и функционал библиотеки NLTK	36	18	8	10	18	зачет	

2.4	POS-маркировка и извлечение именованных сущностей	36	18	8	10	18	зачет	
2.5	Методы тематического моделирования	36	18	8	10	18	зачет	
2.6	Методы нормализации и токенизации средствами NLTK	72	36	12	24	36	зачет	
3	**Python для анализа данных**	**252**	**126**	**52**	**74**	**126**		ПК 1, ПК 4
3.1	Введение в python	36	18	8	10	18	зачет	
3.2	Типы данных в python	36	18	8	10	18	зачет	
3.3	Библиотеки языка python	36	18	8	10	18	зачет	
3.4	Анализ данных. Большие данные	36	18	8	10	18	зачет	
3.5	Математические методы анализа данных в языке python	36	18	8	10	18	зачет	
3.6	Глубокое обучение в языке python	72	36	12	24	36	зачет	
4.	**Итоговая аттестация – аттестационная работа**	**36**					36	ПК 1-7
	ВСЕГО:	**864**	**414**	**112**	**158**	**414**	**36**	

Для всех видов аудиторных занятий академический час устанавливается продолжительностью 45 минут.

Промежуточный контроль проходит на последнем занятии контактной работы с преподавателем.

6. Календарный учебный график

Календарный учебный график – локальный документ, регламентирующий организацию образовательного процесса при реализации программы дополнительного профессионального образования – программы профессиональной переподготовки.

Календарный учебный график разрабатывается и утверждается на каждую учебную группу.

Образовательный период в данной группе начинается по мере ее комплектования.

Первым днем, первой недели обучения, считать день зачисления слушателей на обучение по данной образовательной программе.

Календарный учебный график
Дополнительная профессиональная программа профессиональной переподготовки
«Архитектор данных с фокусом на машинное обучение» (при работе с большими данными)

Учебные недели/ Наименование учебных курсов, дисциплин (модулей) практик	**1 - 2**	**3 - 4**	**5 - 6**	**7 - 8**	**9 -10**	**11 -12**	**13 -14**	**15 -16**	**17 -18**	**19 - 20**	**21 - 22**	**23 - 24**	**25 - 26**	**27 - 28**	**29 - 30**	**31 - 32**	**33 - 34**	**35 - 36**	**37 - 38**	**39**
Фундаментальные алгоритмы машинного обучения	А	А																		
Анатомия алгоритмов обучения		А	А																	
Выбор алгоритма машинного обучения			А	А																
Нейронные сети и глубокое обучение				А	А															
Подготовка данных для алгоритмов машинного обучения					А	А														
Продвинутые методики машинного обучения						А	А													
Обучение без учителя							А	А												
Методы обработки естественного языка								А	А											
Основные библиотеки для обработки естественного языка и их возможности									А	А										
Основные возможности и функционал библиотеки NLTK										А	А									
POS-маркировка и извлечение именованных сущностей											А	А								
Методы тематического моделирования												А	А							
Методы нормализации и токенизации средствами NLTK													А	А						
Введение в python														А	А					
Типы данных в python															А	А				
Библиотеки языка python																А	А			
Анализ данных. Большие данные																	А	А		
Математические методы анализа данных в языке python																		А	А	
Глубокое обучение в языке python																			А	А
Итоговая аттестация																				ИА

Условные обозначения:
А – Аудиторное занятие (лекция, практическое занятие, самостоятельная работа)

ИА - Итоговая аттестация

7. Содержание программ дисциплин (рабочие программы учебных курсов, дисциплин (модулей) практик)

Содержание образовательной программы:

1. Методы машинного обучения

1.1. Фундаментальные алгоритмы машинного обучения.

В данной теме объясняются понятия машинного обучения (ML) и искусственного интеллекта (AI). Дается объяснение понятиям дедуктивного и индуктивного искусственного интеллекта. В рамках материала лекции описываются основные языки машинного обучения, приводятся фундаментальные алгоритмы машинного обучения и методологии описания бизнес-процессов машинного обучения.

Машинное обучение — это подраздел исследований искусственного интеллекта, которое также является подразделом компьютерных наук. Увеличение объема данных и вычислительной мощности также увеличило потребность в обработке данных и получении полезных результатов путем выяснения взаимосвязей между данными. Новое направление называется наукой о данных или аналитикой в целом. Иногда, в зависимости от приложения, платформа, на которой оно работает, может называться аналитикой больших данных. С другой стороны, наука о данных опирается на 2 основных столпа: машинное обучение и статистика.

Тенденция науки о данных также привлекла внимание к исследованиям машинного обучения, которое является одной из двух основных дисциплин, а другая - статистикой. Машинное обучение - относительно старая тема. Некоторые исследователи могут даже найти его корни в древнегреческом, где люди пытаются найти общую формулу для решения всех проблем, или голема из раннего иудаизма, который очень похож на роботов в нашем современном понимании.

Существует три типа алгоритмов машинного обучения.

1. Контролируемое обучение: Регрессия; Классификация;
2. Обучение без учителя: Кластеризация; Уменьшение размерности;
3. Обучение с подкреплением

1.2. Анатомия алгоритмов обучения.

В данной лекции речь пойдет о некоторых практических особенностях, отличающих один алгоритм обучения от другого. Вы уже знаете, что разные алгоритмы обучения могут иметь разные гиперпараметры (C в SVM, ϵ и d в ID3). Алгоритмы оптимизации, такие как градиентный спуск, тоже могут иметь гиперпараметры, например α. Некоторые алгоритмы,

такие как обучение дерева решений, могут принимать качественные признаки. Например, если в наборе данных имеется признак «цвет», принимающий такие значения, как «красный», «желтый» или «зеленый», вы можете оставить его как есть. Но SVM, логистическая и линейная регрессия, а также kNN (с метриками косинусного сходства или евклидова расстояния) ожидают, что все признаки будут иметь числовые значения. Все алгоритмы, реализованные в scikit-learn, работают только с числовыми признаками.

Некоторые алгоритмы, такие как SVM, позволяют аналитикам добавлять весовые коэффициенты для каждого класса. Эти весовые коэффициенты влияют на определение границы решения. Если какой-либо класс имеет большой вес, алгоритм обучения постарается не допускать ошибок в классификации обучающих данных, принадлежащих этому классу (обычно за счет ошибок в других данных). Это может быть важно, если представители какого-то класса находятся в меньшинстве в обучающих данных и вам нужно избежать, насколько это возможно, неправильной их классификации.

Некоторые модели классификации, такие как SVM и kNN, выводят для заданного вектора признаков только класс. Другие, такие как логистическая регрессия или деревья решений, также могут возвращать оценку от 0 до 1, которую можно интерпретировать как степень уверенности модели в прогнозе или вероятность, что входной образец принадлежит определенному классу 1.

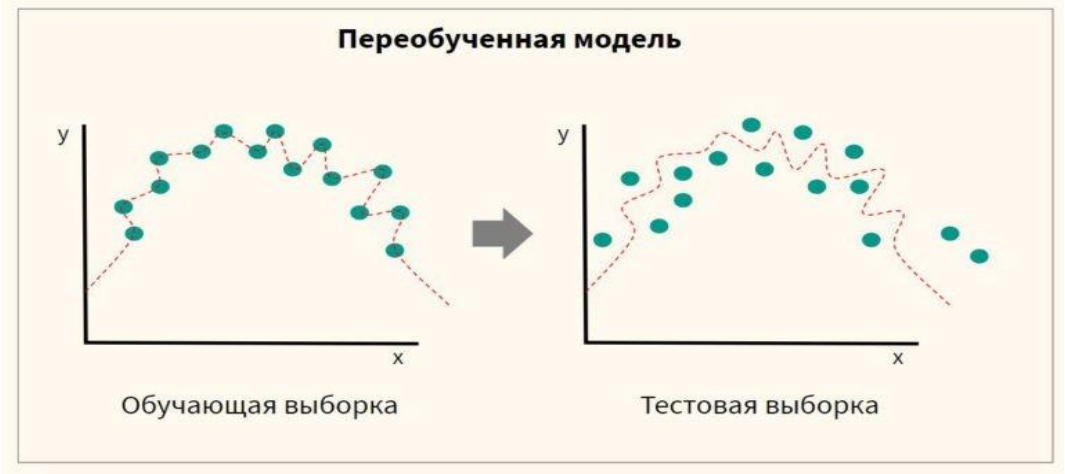
Некоторые алгоритмы классификации (например, обучения дерева решений, логистической регрессии или SVM) строят модель, используя сразу весь набор данных. При появлении дополнительных размеченных данных вам придется пересобрать модель заново. Другие алгоритмы (такие как наивный байесовский классификатор, многослойный персептрон, SGDClassifier/SGDRegressor, PassiveAggressiveClassifier/PassiveAggressiveRegressor в scikit-learn) могут обучаться итеративно, принимая обучающие данные пакетами. При появлении новых обучающих данных вы можете обновить модель, использовав только новые данные.

Наконец, некоторые алгоритмы, такие как обучение дерева решений, SVM и kNN, могут использоваться как для классификации, так и для регрессии, в то время как другие способны решать только задачи одного вида: либо классификацию, либо регрессию, но не обе. Обычно каждая библиотека сопровождается документацией, где объясняется, какую задачу решает каждый алгоритм, какие входные значения допустимы и что возвращает модель. Документация также содержит информацию о гиперпараметрах.

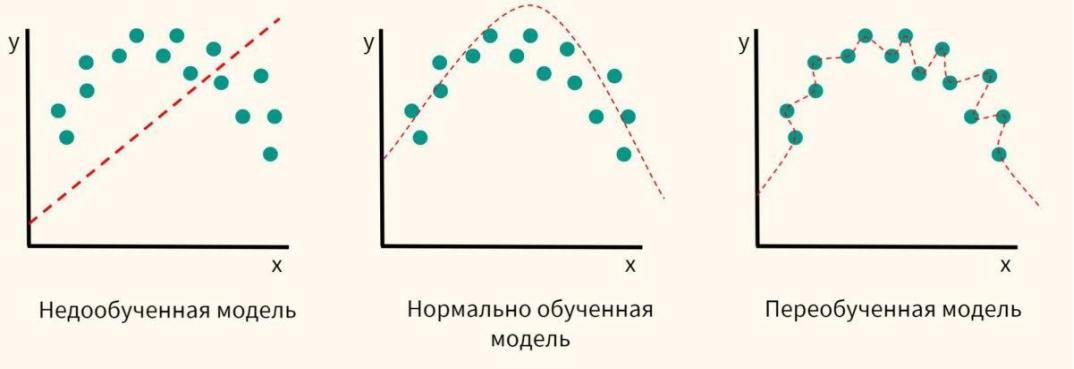
1.3. Выбор алгоритма машинного обучения.

Переобучение (overfitting) случается, когда модель показывает хороший результат на обучающей выборке и сильно «проседает» на тестовых данных.

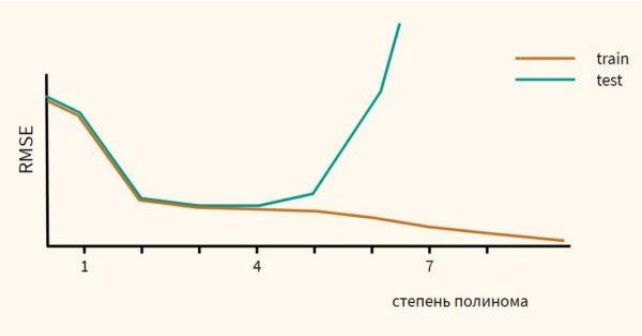
При **недообучении** (underfitting) наоборот у алгоритма не получается уловить зависимость на обучающих данных.



Дополним его для наглядности иллюстрацией того, как может повести себя переобученная модель на тестовых данных.



С точки зрения уровня ошибки на обучающих и тестовых данных можно привести следующий график. Предположим, что мы взяли полиномиальную регрессию и смотрим на значение функции потерь на обучающих и тестовых данных для различных степеней регрессии.



Как вы видите, после определенной границы сложности модели ошибка на тестовых данных начала расти.

1.4. Нейронные сети и глубокое обучение.

Нейронные сети представляют собой семейство моделей, которое начиная с 2012-го постепенно добивается превосходства во всё новых и новых приложениях, во многих став де-факто стандартом. Они были придуманы ещё в 70-х, но техническая возможность и понимание того, как обучать нейросети большого размера, появились лишь примерно к 2011 году, и это дало мощный толчок к их развитию. Совокупность нейросетевых подходов и сама наука о нейросетях носит название глубинного обучения или deep learning. Во многом глубинное обучение основано на двух идеях. Во-первых, это стремление к переходу от построения сложных пайплайнов, каждая компонента которых тренируется сама по себе решать кусочек задачи, к end-to-end обучению всей системы, как одного целого. Так, мы можем обучать не каждый слой отдельно, а все вместе, и не учить какие-нибудь линейные модели поверх случайных лесов, а работать с одной цельной моделью. Во-вторых, это обучение представлений объектов – информативных признаковых описаний, учитывающих структуру данных, построенных лишь на основе самих данных, зачастую неразмеченных. Это позволяет автоматизировать процесс отбора признаков: теперь вместо того, чтобы руками экспертов искать «более информативное или более простое» признаковое описание наших объектов, мы можем получить их автоматически, притом используя не только данные, доступные нам для задачи, но и, к примеру, все тексты мира.

Искусственная нейронная сеть (далее — нейронная сеть) — это сложная дифференцируемая функция, задающая отображение из исходного признакового пространства в пространство ответов, все параметры которой могут настраиваться одновременно и взаимосвязанно (то есть сеть может обучаться end-to-end). В частном (и наиболее частом) случае представляет собой последовательность (дифференцируемых) параметрических преобразований.

Сложную функцию удобно представлять в виде суперпозиции простых, и нейронные сети обычно предстают перед программистом в виде конструктора, состоящего из более-менее простых блоков (слоёв, layers).

Даже самые сложные нейронные сети обычно собираются из относительно простых блоков, подобных этим. Таким образом, их можно представить в виде **вычислительного графа** (**computational graph**), где вершинам промежуточным соответствуют преобразования.

1.5. Подготовка данных для алгоритмов машинного обучения.

Подготовка данных для машинного обучения включает в себя предварительную обработку.

Предварительная обработка и очистка данных должны проводиться до того, как набор данных будет использоваться для обучения модели. Необработанные данные зачастую искажены и ненадежны, и в них могут быть пропущены значения. Использование таких данных при моделировании может приводить к неверным результатам. Эти задачи являются частью процесса обработки и анализа данных группы и обычно подразумевают первоначальное изучение набора данных, используемого для определения и планирования необходимой предварительной обработки.

Реальные данные собираются для последующей обработки из разных источников и процессов. Они могут содержать ошибки и повреждения, негативно влияющие на качество набора данных. Вот какими могут быть типичные проблемы с качеством данных:

- Неполнота: данные не содержат атрибутов, или в них пропущены значения.
- Шум: данные содержат ошибочные записи или выбросы.
- Несогласованность: данные содержат конфликтующие между собой записи или расхождения.

Качественные данные — это необходимое условие для создания качественных моделей прогнозирования. Чтобы избежать появления ситуации «мусор на входе, мусор на выходе» и повысить качество данных и, как следствие, эффективность модели, необходимо провести мониторинг работоспособности данных, как можно раньше обнаружить проблемы и решить, какие действия по предварительной обработке и очистке данных необходимы.

1.6. Продвинутые методики машинного обучения.

В обучении с подкреплением используется способ положительной награды за правильное действие и отрицательной за неправильное. Таким образом, метод присваивает положительные значения желаемым действиям, чтобы побудить агента, и отрицательные значения – к нежелательным. Это программирует нашего агента на поиск долгосрочного и максимального общего вознаграждения для достижения оптимального решения. Эти долгосрочные цели не дают агенту возможности останавливаться на достигнутом. Со временем система учится избегать негативных действий и совершает только позитивные.

Обучение на основе взаимодействия с окружающей средой происходит методом проб и ошибок.

Представьте, что у нас есть стол для игры в пинг-понг и две ракетки. Пусть наша цель – такая система, где мяч не должен пропускаться ни одной из ракеток. Когда каждая из ракеток отбивает мяч, наша позитивная награда увеличивается на единицу (+1), соответственно, в случае пропуска мяча агент получает отрицательное вознаграждение (-1).

Следует помнить, что обучение с подкреплением требует больших вычислительных ресурсов и времени, особенно когда пространство для действий у модели велико.

Область обучения с подкреплением состоит из нескольких алгоритмов, использующих разные подходы. Различия в основном связаны с их стратегиями взаимодействия с окружающей средой:

- State-Action-Reward-State-Action (SARSA). Этот алгоритм обучения с подкреплением начинается с предоставления агенту такого коэффициента, как политика (on-policy). В данном случае политика – это вероятность, с помощью которой алгоритм оценивает шансы определенных действий, приводящих к вознаграждениям или положительным состояниям.
- Q-Learning. В этом подходе к Reinforcement Learning используется противоположный подход. Агент не получает политики (on-policy), соответственно, его исследование окружающей среды является более самостоятельным. В Q-learning у нас нет ограничений на выбор действия (action) для алгоритма. Он полагает, что все последующие выборы actions будут оптимальными по умолчанию, поэтому алгоритм производит операцию выбора исходя из максимизации оценки Q.
- Deep Q-Networks (Глубокие Q-сети). Этот алгоритм использует нейронные сети в дополнение к методам обучения с подкреплением (reinforcement learning). Нейросети осуществляют самостоятельное исследование (research) среды обучения с подкреплением для выбора оптимального значения. То, как алгоритм будет себя вести и подбирать значения, основано на выборке прошлых положительных действий, полученных нейронной сетью.

1.7. Обучение без учителя

Алгоритмы обучения без учителя часто используются для выявления скрытых закономерностей в данных, снижения размерности, а также обнаружения аномалий или выбросов. Они оказываются важным инструментом в анализе данных и подготовке признаков для последующих моделей машинного обучения.

Оценка плотности (Density Estimation)

Оценка плотности является ключевым инструментом в статистике и анализе данных. При оценке плотности мы стремимся оценить распределение вероятности данных. Это может помочь нам найти области высокой и низкой плотности вероятности, что в свою очередь позволяет обнаружить скрытые закономерности или проявления в данных.

Методы оценки плотности могут варьироваться от простых, таких как гистограммы, до более сложных, таких как метод парзеновского окна или гауссовские смеси. Гистограммы разбивают данные на несколько интервалов и считают частоту значений в каждом

интервале, тогда как метод парзеновского окна использует окно фиксированного размера вокруг каждой точки данных для оценки плотности, а гауссовские смеси моделируют распределение данных в виде комбинации нескольких гауссовских распределений.

Кластеризация (Clustering)

Кластеризация в анализе данных — это процесс группировки наблюдений в наиболее однородные кластеры. Цель кластеризации заключается в обнаружении внутренних структур в данных, что помогает выявить сходство внутри кластера и различия между кластерами. Это может использоваться для выявления групп похожих товаров, сегментации пользователей, анализа текстов или изображений и многих других областей.

Существует множество алгоритмов для кластеризации данных, включая K-средних, иерархическую кластеризацию, DBSCAN, методы на основе плотности и многие другие. Каждый из них имеет свои достоинства и ограничения, и выбор подходящего метода зависит от специфики данных и требований к результату.

Сокращение размерности (Dimensionality Reduction)

Сокращение размерности — это процесс уменьшения количества переменных или признаков в данных. Это позволяет уменьшить избыточность и шум в данных, а также улучшить производительность моделей машинного обучения за счет устранения мультиколлинеарности и уменьшения общей сложности моделей.

Методы сокращения размерности включают в себя метод главных компонент (PCA), метод t-распределенных стохастических соседей (t-SNE), методы разреженного сэмплирования и другие. PCA, например, ищет главные компоненты в данных, что позволяет сократить размерность и сохранить наибольшее количество информации. t-SNE, с другой стороны, используется для визуализации данных в двумерном или трехмерном пространстве.

2. Семантическая обработка естественно-языковых текстов

2.1. Методы обработки естественного языка

Обработка естественного языка (NLP, Natural Language Processing) представляет собой область искусственного интеллекта, занимающуюся взаимодействием компьютеров с текстами и речью на естественных языках. Основная цель NLP – создание алгоритмов и систем, способных понимать, интерпретировать и генерировать человеческий язык.

Область применения NLP обширна: от автоматической обработки текстовых данных до создания голосовых помощников и систем машинного перевода. Примеры использования включают чат-ботов, системы поиска информации, анализ отзывов и создание рекомендаций. Сегодня мы подробно рассмотрим ключевые методы и подходы

NLP, акцентируя внимание на обработке текстов, извлечении информации и анализе данных.

2.2. Основные библиотеки для обработки естественного языка и их возможности.

Natural Language Toolkit (NLTK) – это одна из первых и самых известных библиотек для обработки естественного языка (NLP) в Python. Она была создана с целью облегчить исследовательскую и образовательную деятельность в области NLP, предоставляя широкий набор инструментов для работы с текстовыми данными. NLTK включает в себя методы токенизации, стемминга, лемматизации, тегирования частей речи, синтаксического анализа, работы с текстовыми корпусами и многое другое. NLTK превосходит многие библиотеки по широте охвата задач и количеству инструментов. Однако современные библиотеки, такие как SpaCy, предоставляют более производительные решения для задач промышленного уровня. Например, SpaCy лучше подходит для быстрого анализа больших объемов данных и предоставляет модели, готовые к применению в реальном времени.

SpaCy – это одна из самых популярных и производительных библиотек для обработки естественного языка (NLP). Она была создана с упором на простоту использования, высокую производительность и интеграцию в производственные системы. SpaCy предоставляет инструменты для выполнения всех основных задач NLP, таких как токенизация, анализ синтаксиса, извлечение сущностей, лемматизация и многое другое.

Библиотека Stanza – это мощный инструмент для обработки естественного языка (NLP), разработанный командой Стэнфордского университета. Она позволяет решать широкий спектр задач NLP, включая токенизацию, лемматизацию, морфологический и синтаксический анализ, Named Entity Recognition (NER), и многое другое. Stanza предоставляет универсальный интерфейс для обработки текстов на множестве языков, что делает её удобным выбором для исследователей и разработчиков.

2.3. Основные возможности и функционал библиотеки NLTK.

Библиотека NLTK (Natural Language Toolkit) является одной из наиболее известных и широко используемых библиотек для обработки естественного языка. Она предоставляет мощные инструменты для работы со словами, фразами, предложениями и корпусами текстов, что делает её незаменимым инструментом как для обучения, так и для научных исследований. В данном разделе мы подробно рассмотрим возможности работы с текстовыми единицами: словами, фразами, предложениями и корпусами.

NLTK – это универсальная библиотека для обработки текста, предоставляющая инструменты для токенизации, лемматизации, POS-теггинга, анализа фраз и работы с

корпусами. Она поддерживает лингвистический анализ на уровне слов, фраз и предложений, что делает её полезной для образовательных и исследовательских целей. Встроенные корпуса и частотный анализ позволяют изучать текстовые данные на глубоком уровне. NLTK особенно подходит для начального изучения NLP и решения базовых задач обработки естественного языка.

2.4. POS-маркировка и извлечение именованных сущностей.

POS-разметка (от английского Part of Speech tagging) – это процесс определения и присвоения каждому слову в тексте его части речи (существительное, глагол, прилагательное и т.д.) на основе его контекста и значения. Этот процесс является важным этапом обработки естественного языка (NLP), так как позволяет анализировать синтаксическую и семантическую структуру текста, что особенно полезно для задач, таких как машинный перевод, извлечение именованных сущностей и анализ текста.

POS-разметка опирается на грамматические правила и статистику. Для выполнения этой задачи используются как заранее созданные словари с метками, так и алгоритмы машинного обучения. Популярные библиотеки для POS-теггинга в Python – это NLTK (Natural Language Toolkit) и spaCy. Обе библиотеки предоставляют мощные инструменты для работы с текстом и анализа частей речи.

Теггирование слов текста (или POS-теггинг – Part of Speech tagging) представляет собой процесс анализа текста, при котором каждому слову присваивается определенная грамматическая категория, такая как существительное, глагол, прилагательное и т. д. Эта операция является ключевым шагом обработки естественного языка (NLP), поскольку помогает структурировать текст для дальнейшего анализа.

Теггирование текстов не ограничивается только английским языком. Современные библиотеки, такие как Stanza, поддерживают множество языков, включая русский.

POS-маркировка и извлечение именованных сущностей являются фундаментальными задачами обработки естественного языка. POS-маркировка позволяет автоматически определять грамматические категории слов в тексте, что важно для синтаксического и семантического анализа. Использование библиотек, таких как NLTK, spaCy и Stanza, значительно упрощает процесс теггирования слов текста, включая поддержку различных языков и кастомизацию под специфические задачи. Извлечение именованных сущностей (NER) базируется на результатах POS-теггинга и позволяет выделять в тексте значимые объекты, такие как имена, организации и даты. Современные инструменты обеспечивают высокую точность анализа, что делает эти методы незаменимыми в приложениях, связанных с анализом текстов, поисковыми системами и искусственным интеллектом.

2.5. Методы тематического моделирования.

Тематическое моделирование — это метод анализа текстовых данных, который позволяет автоматически обнаруживать скрытые темы в больших объемах текстов. Эти темы представляют собой группы слов, которые часто встречаются вместе и отражают основные идеи или направления содержания документов. Тематическое моделирование применяется в задачах обработки естественного языка (NLP), анализа больших данных, информационного поиска и в исследованиях рынка.

Темы в текстах — это совокупности слов, которые связаны друг с другом на основе их частоты и контекста. Например, в новостных статьях на тему технологий слова «искусственный интеллект», «машинное обучение» и «алгоритмы» могут часто встречаться вместе и, таким образом, формировать одну тему.

Тематическое моделирование помогает не только выявить такие темы, но и определить, какие из них преобладают в каждом отдельном тексте. Это полезно, например, для анализа пользовательских отзывов, новостных публикаций или научных статей.

Латентное размещение Дирихле (LDA) — это популярный метод тематического моделирования, который используется для выявления скрытых тем в текстовых данных. В основе LDA лежит предположение, что каждый документ можно представить как смесь тем, а каждая тема — как распределение слов. Этот метод применяется для анализа больших текстовых корпусов, например, новостных статей, научных публикаций или пользовательских отзывов.

Методы тематического моделирования, такие как LSA, LDA и NMF, позволяют выявлять скрытые темы в текстах и анализировать большие объемы данных. Латентный семантический анализ (LSA) использует матричное разложение для обнаружения связей между словами и документами, упрощая анализ текстовой информации. Латентное размещение Дирихле (LDA) применяет вероятностный подход, позволяя интерпретировать тексты как смеси тем, что делает его подходящим для анализа сложных и многозначных данных. Эти методы широко применяются в информационном поиске, анализе отзывов и классификации текстов. Современные модели, такие как BERTopic, дополняют классические подходы, обеспечивая более глубокий и точный анализ.

2.6. Методы нормализации и токенизации средствами NLTK.

NLTK (Natural Language Toolkit) предоставляет широкий спектр инструментов для анализа текстов, от простых операций, таких как токенизация и нормализация, до сложных задач, включая POS-теггинг, построение синтаксических деревьев и обучение языковым моделям.

NLTK используется как начинающими, так и опытными специалистами в области NLP благодаря своей простоте, обширной документации и встроенным наборам данных. Ниже рассмотрим основные возможности библиотеки.

Выделение именованных сущностей (Named Entity Recognition, NER) — это одна из ключевых задач в области обработки естественного языка (NLP), направленная на извлечение из текста информации о сущностях, таких как имена людей, географические объекты, организации, даты, числовые значения и другие важные элементы. Задача NER является основой для многих приложений в области анализа текста, таких как автоматический анализ новостей, извлечение информации, создание чат-ботов, а также в сфере информационного поиска и создания рекомендательных систем.

Нормализация текста — это процесс преобразования текста в стандартную форму, что позволяет уменьшить разнообразие представлений данных и привести слова к единой форме. Нормализация необходима для того, чтобы можно было эффективно анализировать и сравнивать слова, независимо от их различных вариантов написания.

Методы нормализации и токенизации являются основными этапами обработки текста в области обработки естественного языка (NLP). Токенизация с помощью библиотеки NLTK позволяет разбить текст на составляющие элементы, такие как слова или предложения, что является важным для последующего анализа. Нормализация включает в себя операции, такие как приведение слов к единой форме с помощью стемминга и лемматизации, что помогает устранить избыточность и привести данные к удобному виду для анализа. Стемминг и лемматизация — два различных подхода к нормализации слов, где стемминг быстрее, но менее точен, в то время как лемматизация предоставляет более точные результаты, но требует дополнительных ресурсов. Использование этих методов в связке с другими инструментами NLTK помогает эффективно решать задачи, такие как анализ текстов, извлечение информации и построение языковых моделей.

3. Python для анализа данных

3.1. Введение в python.

Python — это мощный, гибкий и интуитивно понятный язык программирования, который идеально подходит как для новичков, так и для опытных разработчиков. Благодаря своей простоте и лаконичности синтаксиса, Python позволяет быстро осваивать основы программирования и переходить к решению реальных задач.

Рассмотри ключевые аспекты Python:

- Установку и настройку среды разработки (Anaconda, Spyder, Jupyter Notebook).
- Запуск тестовой программы и выполнение базовых операций.

- Основные элементы языка: ввод-вывод данных, переменные, циклы, условные конструкции и функции.

Однако, это только начало пути в изучении Python. Для более глубокого понимания языка рекомендуется изучить:

- Структуры данных (списки, кортежи, множества, словари).
- Объектно-ориентированное программирование (ООП) (классы, объекты, наследование, полиморфизм).
- Работу с файлами (чтение и запись файлов, обработка исключений).
- Использование стандартных и сторонних библиотек (NumPy, Pandas, Matplotlib, TensorFlow и другие).
- Алгоритмы и структуры данных (поиск, сортировка, рекурсия, деревья, графы).
- Разработку веб-приложений (Django, Flask).
- Анализ данных и машинное обучение (Scikit-learn, PyTorch, OpenCV).

Python уже активно применяется в различных сферах: от веб-разработки до искусственного интеллекта и автоматизации.

3.2. Типы данных в python.

Переменные в Python предназначены для хранения данных различных типов. Они позволяют присваивать значения и обращаться к ним по имени в процессе выполнения программы. Переменная хранит данные одного из типов данных. В Python существует множество различных типов данных. Python является языком с динамической типизацией - переменная не привязана к определенному типу.

Тип переменной определяется исходя из значения, которое ей присвоено. Так, при присвоении строки в двойных или одинарных кавычках переменная имеет тип str. При присвоении целого числа Python автоматически определяет тип переменной как int. Чтобы определить переменную как объект float, ей присваивается дробное число, в котором разделителем целой и дробной части является точка.

Целые числа представляют собой неизменяемый числовой тип данных, используемый для хранения целочисленных значений произвольной величины. В Python тип int автоматически определяется при присвоении числового значения без десятичной точки и поддерживает как положительные, так и отрицательные числа, включая ноль.

В Python строки обрамляются либо одинарными, либо двойными кавычками — разницы между ними нет. Например, ‘Hello’ и "Hello" абсолютно идентичны. Если нужно использовать кавычки внутри строки, можно комбинировать их виды: например, '"вложенные кавычки"' или '"вот так"'.

Словари в Python предназначены для хранения данных в виде пар ключ:значение (key:value). Они представляют собой упорядоченные (начиная с Python 3.7) и изменяемые коллекции, которые не допускают дублирования ключей. Словари записываются в фигурных скобках {}, где каждый элемент состоит из ключа и соответствующего ему значения.

Кортежи представляют собой неизменяемые упорядоченные коллекции для хранения данных. В отличие от списков, которые можно модифицировать после создания, кортежи остаются постоянными - их элементы нельзя изменить, добавить или удалить после создания. Это делает их особенно полезными для хранения константных данных, которые не должны изменяться в ходе выполнения программы.

Списки представляют собой изменяемые упорядоченные коллекции, которые позволяют хранить набор элементов в одной переменной. Они создаются с помощью квадратных скобок и могут содержать элементы разных типов данных, включая числа, строки, другие списки или любые другие объекты Python. В отличие от кортежей, списки являются изменяемыми, что означает возможность добавления, удаления или изменения элементов после создания списка.

3.3. Библиотеки языка python.

Python – это мощный язык программирования, который получил широкое распространение благодаря своей простоте и богатой экосистеме библиотек. В данной лекции мы рассмотрим основные аспекты работы с библиотеками Python, включая их установку, использование, а также разберем наиболее популярные библиотеки для анализа данных.

Python широко используется в анализе данных благодаря наличию специализированных библиотек. Python позволяет импортировать библиотеки различными способами. После импорта библиотек можно использовать их функции.

Python предлагает мощный инструментарий для анализа данных и научных вычислений. В данной лекции мы рассмотрели основы работы с библиотеками: установку через PIP, базовые библиотеки для анализа данных, принципы импорта и примеры использования библиотек. Знание этих инструментов позволит эффективно решать задачи анализа данных и визуализации.

3.4. Анализ данных. Большие данные.

Мир сегодня генерирует огромное количество данных, и способность эффективно их анализировать стала важнейшим навыком для специалистов в самых разных областях – от бизнеса и маркетинга до науки и медицины. Анализ данных позволяет извлекать полезные сведения из сырых наборов информации, предсказывать будущие события и

делать обоснованные выводы. Для успешного проведения анализа необходимы специальные инструменты и методы, которые помогут справиться с потоками больших данных.

В этом контексте ключевыми являются библиотеки Python, такие как NumPy, pandas, Matplotlib и Seaborn, каждая из которых решает конкретные задачи в процессе работы с данными. От быстрой векторизации операций через NumPy до удобной манипуляции с таблицами в pandas и построения наглядных графиков – эти инструменты обеспечивают мощный арсенал для извлечения знаний из больших объемов информации.

Обучение основам работы с этими библиотеками открывает путь к пониманию и применению методов анализа данных, что позволит вам уверенно справляться с различными задачами, начиная от предварительной очистки данных и заканчивая построением сложных моделей и визуализацией результатов.

Процессы чтения, очистки и визуализации данных являются фундаментальной основой для любого современного анализа данных. Эти этапы определяют качество итогового результата и точность выводов, сделанных на основании собранной информации. Чтение данных из различных источников – будь то файлы, базы данных или API – закладывает основу для всей последующей работы. Правильное форматирование и проверка целостности данных обеспечивают надежную базу для дальнейших шагов.

Очистка данных – ключевой этап, который требует особого внимания. Недостоверные, неполные или зашумленные данные могут исказить результаты анализа и привести к ложным выводам. Поэтому процедуры устранения пропусков, нормализации и коррекции ошибок являются обязательными для достижения качественных результатов.

Визуализация данных служит мостом между числовыми показателями и человеческим восприятием. Грамотно подобранные графики и диаграммы способны раскрыть скрытые закономерности, которые трудно заметить при работе с сырыми цифрами. Они облегчают понимание и интерпретацию данных, делая их доступными для широкой аудитории.

Таким образом, качественное выполнение этих этапов – залог успешных исследовательских проектов, достоверных прогнозов и правильных управленческих решений. Без должного внимания к каждому из них невозможно обеспечить надежность и воспроизводимость результатов.

3.5.Математические методы анализа данных в языке python.

Вероятностные модели играют ключевую роль в науке, технике и повседневной жизни, помогая оценивать неопределенности и принимать обоснованные решения. В рамках теории вероятностей особую значимость приобретают понятия случайных величин

и правила Байеса, которые служат фундаментом для широкого спектра прикладных задач – от статистики и анализа данных до машинного обучения и принятия решений.

Случайная величина –это математическая модель, используемая для описания неопределенного результата эксперимента или наблюдения. В отличие от детерминированных величин, случайные величины принимают значения в соответствии с некоторым распределением вероятности. Простейшими примерами случайных величин могут служить бросок монеты (где результатом является "орел" или "решка") или измерение высоты человека (где результат может варьироваться в пределах некоторого диапазона).

Градиентный спуск остается краеугольным камнем в мире оптимизации и машинного обучения, и его важность вряд ли уменьшится в будущем. Освоение этого метода открывает широкие возможности для решения самых разнообразных задач, от разработки нейронных сетей до оптимизации производственных процессов.

3.6.Глубокое обучение в языке python.

Машинное обучение (ML) — это подраздел искусственного интеллекта, основанный на использовании алгоритмов, которые автоматически улучшают свою производительность через опыт (данные).

Глубокое обучение (DL) — подкласс ML, использующий многослойные нейронные сети для автоматического извлечения признаков из сырых данных.

Обучение с учителем (Supervised Learning). Модель обучается на размеченных данных, где каждому примеру соответствует правильный ответ. Dataset обычно представляется в виде таблицы, где строки соответствуют записям (объектам), а столбцы — признакам (характеристикам) этих объектов. Данные в сетах могут быть разных типов, и правильное понимание этих типов критично для их анализа и обработки.

Обучение без учителя (Unsupervised Learning). Модель ищет скрытые паттерны в данных без готовых ответов. Такая модель работает с неразмеченными данными (только X). Она выявляет структуру данных. Самые распространенные алгоритмы: K-means, Иерархическая, PCA (метод главных компонент), кластеризация.

Обучение с подкреплением (Reinforcement Learning). Агент обучается через взаимодействие со средой, получая награды за правильные действия. Как и в случае с обучением без учителя, здесь нет готового датасета. Обучение проходит через систему проб и ошибок.

Искусственный нейрон представляет собой базовую вычислительную единицу нейронной сети, моделирующую принципы работы биологического нейрона. Математически нейрон можно представить как функцию, принимающую вектор входных

сигналов x = (x_1, x_2, ..., x_n) и возвращающую скалярное выходное значение a. Вычисление происходит в два этапа: сначала вычисляется взвешенная сумма входных сигналов $z = \Sigma(w_i * x_i) + b$, где w_i - весовые коэффициенты, b - смещение (bias), затем применяется нелинейная функция активации a = f(z).

Алгоритм обратного распространения ошибки (Backpropagation). Алгоритм обратного распространения ошибки представляет собой фундаментальный метод обучения многослойных нейронных сетей, основанный на цепном правиле дифференцирования сложных функций.

Batch Normalization помогает стабилизировать обучение, устраняя проблему «внутреннего ковариационного сдвига», которая возникает, когда изменяются распределения входных данных слоев в процессе обучения. BN нормализует входные данные каждого слоя, что ускоряет и стабилизирует обучение.

Промежуточная аттестация: вопросы для подготовки

1. Какой оператор используется для целочисленного деления?
2. Что выведет print(type(5 / 2))?
3. Как создать кортеж из одного элемента?
4. Какой метод Pandas позволяет выполнить "ленивую загрузку" данных при работе с большими файлами?
5. Какая функция в NumPy позволяет создать массив с числами, равномерно распределенными в логарифмическом масштабе?
6. Какой тип данных в NumPy позволяет эффективно хранить гетерогенные данные (например, строки и числа)?
7. Как вычислить стандартное отклонение массива?
8. Как добавить слой Dropout?
9. Какой метод оптимизации использует моменты?
10. Как реализовать callback в Keras?
11. Как уменьшить переобучение?
12. Как токенизировать текст в NLTK?
13. Как создать TF-IDF вектор?
14. Как закодировать текст в числа?
15. Какой алгоритм используется для классификации текста?
16. Как удалить стоп-слова?
17. Как называется процесс разделения текста на отдельные предложения?

18. В какой библиотеке реализован метод POS-тэггинга (Part-of-Speech tagging)?
19. Какую функцию использует NLTK для работы с корпусом Brown?
20. В чем разница между униграммой и биграммой?
21. Какое значение принимает параметр n в n-граммах?
22. Какая библиотека предоставляет самый быстрый алгоритм для работы с зависимостями синтаксического разбора?
23. Как использовать NLTK для построения дерева синтаксического разбора?
24. Какова роль скрытого марковского процесса в обработке естественного языка?
25. Какие научные дисциплины объединяет в себе технология NLP (Обработка естественного языка)?
26. Каковы основные этапы разработки системы NLP для задачи классификации текстов?
27. Что такое Active Learning и как оно может быть полезно в NLP?
28. Что такое Attention Mechanism и как он используется в моделях NLP?
29. Что такое Text Clustering?
30. Что такое векторизация текстовых данных?
31. Что такое дедупликация?
32. Что такое морфологический анализ текста?
33. Что такое синтаксический анализ текста?
34. Какой параметр определяет размер шага улучшения, выполняемого на каждой итерации градиентного спуска?
35. Какой наиболее важный этап в генетическом алгоритме?
36. Выделите модель, которая объединяет прогнозы для получения точного прогноза?
37. Определите из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?
38. Определите, для какого типа задач подходит алгоритм линейной регрессии?
39. Определите метод регуляризации для борьбы с мульти коллинеарностью признаков?
40. Определите технологию, которая может работать как со структурированными, так и с неструктурированными типами данных?
41. Какие нейронные сети подходят для работы с изображениями?
42. Сколько типов обучения доступно в машинном обучении?
43. Какая наиболее распространенная проблема при использовании машинного обучения?

44. Вы прогнозируете, является ли электронное письмо спамом. Основываясь на характеристиках, вы получили оценку вероятности 0,75. Определите, что означает эта предполагаемая вероятность?
45. Определите, какова функция затрат логистической регрессии?
46. Вы прогнозируете, является ли электронное письмо спамом. Основываясь на характеристиках, вы получили оценку вероятности 0,75. Определите, что означает эта предполагаемая вероятность? Пороговое значение для различения классов составляет 0,5.
47. Выделите классические методы обучения с учителем?
48. Выделите методы машинного обучения?
49. Определите два основных типа переменных в статистике?
50. Алгоритм анализа главных компонентов – это?

Итоговая аттестация: темы выпускных квалификационных работ

1. Анализ рентгенограмм в диагностике стоматологических заболеваний на основе методов машинного обучения
2. Исследование методов машинного обучения в задаче автоматической классификации изображений
3. Применение искусственного интеллекта для интеграционных решений в ИТ-проектах
4. Исследование и оптимизация бизнес процессов компании, занимающаяся предоставлением облачных решений для бизнеса
5. Анализ языковой личности с использованием искусственного интеллекта на базе русского, английского и китайского языков для определения психоэмоционального состояния индивидуума
6. Исследование процессов обнаружения инфекционных заболеваний на основе анализа рентгенограмм легких
7. Анализ рентгенограмм в диагностике стоматологических заболеваний на основе методов машинного обучения
8. Исследование методов машинного обучения в задаче автоматической классификации изображений
9. Применение искусственного интеллекта для интеграционных решений в ИТ-проектах
10. Исследование и оптимизация бизнес процессов компании, занимающаяся предоставлением облачных решений для бизнеса

11. Использование систем искусственного интеллекта в криминалистике.
12. Анализ тональности текста сообщений социальных сетей.
13. Системы искусственного интеллекта в распознавании образов (на примере).
14. Исследование процессов проектирования и разработки систем искусственного интеллекта.
15. Исследование и анализ применения статистических методов при исследовании научных статей.
16. Исследование проблем разработки и внедрения интеллектуальных систем, основанных на нейронных сетях.
17. Разработка концепции построения информационной системы с использованием методов искусственного интеллекта (на примере).
18. Исследование и анализ форм и методов отображения и представления динамики деятельности компании за длительный период.
19. Использование методов машинного обучения для анализа графических изображений (на примере).
20. Влияние методов искусственного интеллекта на экономику и бизнес (на примере).
21. Инновационное развитие искусственного интеллекта и машинного обучения в современной экономике.
22. Применение искусственного интеллекта в банковской сфере (на примере).
23. Использование методов машинного обучения для установления авторства программы (на примере).
24. Тенденции развития искусственного интеллекта и применение диалоговых интеллектуальных систем с использованием методов машинного обучения.
25. Искусственный интеллект и робототехника (на примере).
26. Экономические аспекты и практические основы развития искусственного интеллекта в России.
27. Искусственный интеллект как дискурс самопознания и самоорганизации цифрового социума.
28. Исследование и анализ возможности использования искусственного интеллекта в образовании.
29. Искусственный интеллект как технологическая инновация для ускоренного развития экономики.
30. Извлечение новых терминов их текстов научных публикаций и построение трендов изменения их значений.

8. Организационно-педагогические условия программы

8.1.Материально-технические условия реализации программы.

Реализация программы профессиональной переподготовки осуществляется на материально-технической базе АНО ВО «Российский новый университет», обеспечивающей проведение всех видов учебных занятий, предусмотренных учебным планом.

Учебный процесс обеспечен учебной аудиторией, соответствующей санитарно-гигиеническим требованиям для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, самостоятельной работы, хранения и профилактического обслуживания учебного оборудования. Помещение укомплектовано мебелью и техническими средствами обучения, служащими для представления учебной информации. Аудитория соответствует нормам освещенности, оснащена системой кондиционирования воздуха.

В учебном помещении имеется необходимая для процесса обучения компьютерная техника, учебно-наглядные пособия, обеспечивающие тематические иллюстрации, соответствующие содержанию программы, освоению дисциплин (модулей).

Помещение подключено к сети "Интернет", также в нем обеспечен доступ в электронную информационно-образовательную среду организации. Рабочее место преподавателя оснащено web-камерой с микрофоном и гарнитурой, необходимой для работы в MS Teams.

8.2. Учебно-методическое и информационное обеспечение программы.

Слушателям предоставляется бесплатный доступ к ресурсам электронной информационно-образовательной среды на сайте Института. Каждый слушатель во время самостоятельной подготовки обеспечивается рабочим местом в компьютерном классе или через выход в Интернет получает доступ к использованию электронных изданий, в соответствии с объемом изучаемых дисциплин.

Каждый слушатель на время занятий обеспечивается комплектом учебно-методических материалов, содержащим электронные и печатные информационные разработки, учебные видеофильмы (тиражируются по требованию).

8.3. Кадровое обеспечение образовательного процесса.

К реализации программы привлечены представители образовательных организаций высшего образования, имеющие высшее образование, ученую степень кандидата/доктора наук, стаж научно-педагогической работы более трех лет.

9. Контроль и оценка результатов освоения программы

9.1. Формы аттестации

Реализация программы профессиональной переподготовки включает в себя промежуточную и итоговую аттестацию.

Текущий контроль проводится по итогам самостоятельной - внеаудиторной работы слушателя.

Промежуточная аттестация проводится по итогам освоения дисциплины (модуля) в форме зачета.

Завершается освоение профессиональной программы переподготовки итоговой аттестацией обучающихся в форме итогового семинара, по результатам которого выставляется оценка.

9.2. Оценочные средства

Критерии оценки уровня овладения компетенциями

Оценка	**Характеристики ответа слушателя**
отлично	• глубоко и всесторонне усвоил программный материал и проявляет способности решать типовые задачи в различных областях профессиональной практики; • уверенно, логично, последовательно и грамотно излагает программный материал; • умело обосновывает и аргументирует выдвигаемые им идеи при решении профессиональных задач и анализе организационно-управленческой и других видов профессиональной деятельности; • делает аргументированные выводы и обобщения.
хорошо	• твердо усвоен программный материал, грамотно и, по существу, излагает его, опираясь на знания основной литературы, владеет умениями решать типовые задачи в отдельных областях профессиональной практики; • не допускает существенных неточностей при изложении учебного материала; • увязывает усвоенные знания с практической деятельностью; • аргументирует научные положения; • делает выводы и обобщения.
удовлетворительно	• усвоен только основной программный материал, по существу, излагает его, опираясь на знания только основной литературы;

	• допускает несущественные ошибки и неточности при изложении учебного материала; • испытывает затруднения в практическом применении полученных знаний; • слабо аргументирует научные положения; • затрудняется в формулировании выводов и обобщений; • в основном владеет системой понятий.

Итоговая аттестация слушателей является обязательной и осуществляется после освоения программы профессиональной переподготовки в полном объеме.