

Общество с ограниченной ответственностью «1Т»

УТВЕРЖДАЮ

Генеральный директор ООО «1Т»

(В.В. Кармаза)

08 » апреля 2024



ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА ПОВЫШЕНИЯ
КВАЛИФИКАЦИИ

«Архитектор в области искусственного интеллекта»

Специальность: Архитектор в области искусственного интеллекта (AI Architect)

Целевое назначение: Разработка и использование технологий искусственного
интеллекта

Срок обучения: 260 академических часов

Москва 2024

ОГЛАВЛЕНИЕ

ОБЩИЕ ДАННЫЕ О ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПРОГРАММЕ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ «АРХИТЕКТОР В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»	3
1 ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ	9
1.1 Актуальность	9
1.2 Категория слушателей. Требования к уровню подготовки слушателя	12
1.3. Область профессиональной деятельности	13
1.4 Планируемые результаты обучения	14
2. УЧЕБНЫЙ (ТЕМАТИЧЕСКИЙ) ПЛАН	22
3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК	26
4. РАБОЧИЕ ПРОГРАММЫ МОДУЛЕЙ УЧЕБНОГО КУРСА ПРОГРАММЫ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ «АРХИТЕКТОР В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»	27
4.1 Рабочая программа модуля 1. Базовый	27
4.2 Рабочая программа модуля 2. Профильный	43
5. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ	61
6. СИСТЕМА ОЦЕНКИ КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ	64

ОБЩИЕ ДАННЫЕ О ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПРОГРАММЕ
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ «АРХИТЕКТОР В ОБЛАСТИ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА»

№	Название	Описание
1.1	Название программы	Архитектор в области искусственного интеллекта
1.2	Цель обучения	Получение слушателями компетенций, необходимых для профессиональной деятельности архитектора в области искусственного интеллекта (далее - ИИ) для разработки и применения технологических решений в области искусственного интеллекта и в смежных областях
1.3	Специальность	Архитектор в области искусственного интеллекта (AI Architect)
1.4	Форма обучения	Очно-заочная форма обучения, осуществляемая с применением электронного обучения и дистанционных образовательных технологий (онлайн-вебинары и т. п.) без отрыва от производства
1.5	Количество академических часов	260
1.6	Количество слушателей, которое может обеспечить обучением провайдер по Образовательной программе по одному потоку в срок до 25 ноября	210
1.7.	Стоимость обучения	80 000

Дополнительная профессиональная программа (программа повышения квалификации) «Архитектор в области искусственного интеллекта» (далее – Программа) разработана:

– в соответствии с нормами Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;

– с учетом постановления Правительства Российской Федерации от 13 мая 2021 г. № 729 «О мерах по реализации программы стратегического лидерства «Приоритет-2030» (в редакции постановления Правительства Российской Федерации от 14 марта 2022 г. № 357 «О внесении изменений в постановление Правительства Российской Федерации от 13 мая 2021 г. № 729»);

– с учетом требований приказа Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», с изменениями, внесенными приказом Минобрнауки России от 15 ноября 2013 г. № 1244 «О внесении изменений в Порядок организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденный приказом Министерства образования и науки Российской Федерации от 1 июля 2013 г. № 499»;

– с учетом приказа Министерства образования и науки РФ от 23 августа 2017 г. № 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»; паспорта федерального проекта «Искусственный интеллект» национальной программы «Цифровая экономика Российской Федерации»;

– с учетом Методических рекомендаций-разъяснений Минобрнауки России по разработке дополнительных профессиональных программ на основе профессиональных стандартов от 22 апреля 2015 года № ВК-1030/06;

– с учетом приказа Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации от 28 февраля 2022 г. № 143 «Об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» и признании утратившими силу некоторых приказов Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации»;

– с учетом федеральных государственных образовательных стандартов (далее вместе – ФГОС ВО) по направлению подготовки 38.03.05 Бизнес-информатика (приказ Минобрнауки России от 29.07.2020 N 838), 09.04.04. Программная инженерия (Приказ Минобрнауки России от 19.09.2017 N 932 (ред. от 08.02.2021));

– на основе анализа требований рынка труда в сфере искусственного интеллекта и анализа данных.

В результате освоения данной программы выпускник программы «Архитектор в области искусственного интеллекта» должен:

Знать:

– функции, задачи, навыки, содержание работы архитектора в области искусственного интеллекта;

- определения, историю развития и главные тренды искусственного интеллекта;
- основные типы, источники больших данных, проблемы безопасности;
- методы и программный инструментарий технологий больших данных;
- синтаксис и семантику основных языков искусственного интеллекта (Python, JAVA/C#/C++/Scala) и основные приемы программирования на них;
- основные алгоритмы, используемые в языках программирования;
- базовые понятия и возможности теории игр для их применения в области машинного обучения и искусственного интеллекта;
- методы, инструментальные средства и стандарты статистического и описательного анализа данных;
- системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, SPARK и др.): принципы работы, функционал, язык запросов, преимущества и недостатки;
- основные платформы данных (облачные и внутрикорпоративные), применяемые при разработке решений на основе ИИ;
- архитектуру, принципы, процесс, стадии и методологии разработки решений на основе ИИ;
- основные виды представления данных: табличные, графовые, временные ряды;
- принципы, содержание и области применения массово-параллельной обработки и анализа данных для ускорения машинного обучения;
- основы методов оптимизации;
- технологии, методы и инструменты машинного обучения для решения профессиональных задач;
- технологии распределенной обработки данных;
- принципы и содержание работы с распределенными кластерными системами;
- основные уровни представления данных;
- технологии и методы потоковой обработки данных;
- архитектуры современных нейронных сетей и их использование для решения профессиональных задач.

Уметь:

- применять системы обработки и анализа больших массивов данных;
- применять системы контроля версий Git для разработки решений на основе языков программирования;
- применять язык SQL для прикладных решений;

- применять язык программирования Python и библиотеки при разработке решений на основе ИИ;
- использовать основные статистические методы анализа данных;
- визуализировать анализируемые данные;
- проводить EDA — разведочный анализ данных;
- моделировать данные в хранилищах (DWH, Data Lake, NoSQL хранилища данных);
- применять методы машинного обучения для работы с большими данными;
- использовать шины данных (kafka) при обработке больших данных;
- строить модель для прогнозирования временного ряда;
- строить модель для рекомендательных систем;
- использовать контейнеризацию и развертывать модель в производство;
- применять технологии компьютерного зрения для обработки изображений и видео, задач биометрии;
- применять технологии обработки естественного языка для анализа и генерации текстов;
- применять технологии обучения с подкреплением для задач автоматизации производства и транспорта;
- использовать цифровые платформы анализа данных для решения профессиональных задач.

Владеть:

- языками программирования (Python, JAVA/C#/C++/Scala) для реализации методов анализа и структурирования данных;
- основными инструментами для работы с данными (Git, Docker, CI/CD, RestAPI, Jupyter notebook, kaggle, Nvidia Cuda, VS, IntelliJIdea);
- навыками работы с основными библиотеками Python для визуализации данных (Matplotlib, Seaborn, Plotly);
- навыком создания логической и физической базы данных;
- навыком установки Hadoop на виртуальную машину;
- навыком использования Docker;
- навыком работы с MapReduce, PySpark;
- навыком BI-аналитики;
- навыком работы с Numpy;
- способами улучшить качество модели с помощью методов Feature engineering;

- навыком использования библиотеки PyTorch, применения GPU в PyTorch;
- навыком построения полносвязной нейронной сети для задачи классификации;
- навыком работы с генеративными моделями GAN для задач генерации;
- навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий;
- навыком работы с технологией RL — обучения с подкреплением;
- языками запросов для обращения к СУБД;
- основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных;
- методами машинного обучения и нейронных сетей;
- математическими методами анализа данных;
- культурой постановки и планирования последовательности решения задач анализа данных и классификации;
- навыками анализа реальных задач из различных предметных областей на уровне отдельных подходов и коллективами алгоритмов;
- навыками работы с библиотеками, программными платформами (фреймворками) и программными комплексами машинного обучения.

Разработчики программы:

Борисов Вадим Владимирович, профессор кафедры вычислительной техники, филиал НИУ «МЭИ» в г. Смоленске, д.т.н., профессор,

Бен Режеб Тауфик Бен Камель, декан факультета «Радио и телевидение» МТУСИ,

Дроздов Сергей Николаевич, доцент Института компьютерных технологий и информационной безопасности, ФГАОУ ВО «Южный федеральный университет», к.т.н.,

Захарова Алена Александровна, главный научный сотрудник ИПУ РАН, д.т.н., доцент,

Ерохин Кирилл Сергеевич, руководитель отдела Data Science ООО «1Т»,

Ерохин Виталий Александрович, ведущий специалист по ИИ, ООО «1Т»,

Кузин Антон Алексеевич, старший преподаватель, Департамент математического и компьютерного моделирования Института математики и компьютерных технологий МТУСИ,

Бондаренко Евгений Владимирович, ведущий специалист по ИИ, ООО «1Т»,

Козлова Анна Сергеевна, специалист по ИИ, ООО «1Т»,

Юров Юрий Аркадьевич, специалист по ИИ, ООО «1Т»,

*Дон Юрий Никитич, разработчик машинного обучения, ООО «Рубиус тех»,
Софьина Каролина Германовна, специалист по ИИ, ООО «1Т».*

1 ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ.

1.1 Актуальность.

Ускоренное внедрение цифровых технологий в экономике и социальной сфере создает условия для высокотехнологичного бизнеса, повышает конкурентоспособность страны на глобальном рынке, укрепляет национальную безопасность и повышает качество жизни людей. В рамках реализации Указа Президента Российской Федерации от 7 мая 2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года», в том числе с целью решения задачи по обеспечению ускоренного внедрения цифровых технологий в экономике и социальной сфере, Правительством Российской Федерации на базе программы «Цифровая экономика Российской Федерации» сформирована национальная программа «Цифровая экономика Российской Федерации», утвержденная протоколом заседания президиума Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам от 4 июня 2019 г. № 7.

Россия должна решить целевые задачи кадровой политики в цифровой экономике, подготовки высококлассных специалистов в информационно-технологической сфере, найти способы и приемы мотивации населения по глубокой интеграции в своей повседневной жизни и профессиональной деятельности объектов цифровизации. Совершенствование системы образования обеспечит подготовку квалифицированных кадров для цифровой экономики, а граждане получат новые возможности и мотивацию для освоения цифровых компетенций.

Под искусственным интеллектом в Программе развития в соответствии с «Национальной стратегией развития искусственного интеллекта на период до 2030 года», утвержденной Указом Президента РФ от 10.10.2019 г. № 409 (далее – Стратегия), понимается комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Для продуктивной реализации задач указанной Стратегии на основе технологий искусственного интеллекта необходимо повышение эффективности процессов планирования, прогнозирования и принятия управленческих решений, использование автономного интеллектуального оборудования и робототехнических комплексов, интеллектуальных систем управления логистикой.

Одним из приоритетных направлений научно-технического развития Российской Федерации согласно Стратегии научно-технологического развития Российской Федерации, утвержденной Указом Президента Российской Федерации от 1 декабря 2016 г. № 642, обозначено создание технологий, являющихся основой инновационного развития внутреннего рынка продуктов и услуг, что может быть осуществлено за счет перехода к цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создания систем обработки больших объемов данных, машинного обучения и искусственного интеллекта.

Для развития рынка ИИ в России в 2020 году был создан федеральный проект «Искусственный интеллект» национального проекта «Цифровая экономика».

Задача федерального проекта «Искусственный интеллект» — создать условия для того, чтобы предприятия и граждане использовали продукты и услуги, основанные на преимущественно отечественных технологиях искусственного интеллекта, обеспечивающих качественно новый уровень эффективности деятельности.

Задачи, на решение которых человек раньше тратил довольно продолжительное время, искусственный интеллект может выполнить за несколько секунд. Уже сегодня с помощью ИИ в десятки раз быстрее открывают банковские счета и проводят закупки, разрабатывают новые лекарства, инвестируют на фондовом рынке и могут с точностью до минут определить время задержки рейса. Искусственный интеллект называют «новым электричеством»: он меняет целые отрасли бизнеса, а в будущем, возможно, изменит и облик всей цивилизации.

Искусственный интеллект способен снижать риски износа и повреждения оборудования, а также создавать систему безопасности для различных компаний. Автоматизация ручного труда также является важной и неоднозначной темой, поскольку использование алгоритмов искусственного интеллекта в промышленности способно вытеснить из этой сферы человеческий труд. Автоматизированные технологии выполняют сложные процессы быстрее и качественнее, чем человек, они способны работать 24 часа в сутки. Следует подчеркнуть, что основная цель внедрения высокоинтеллектуальных решений сегодня – это не полная замена человека в производственных и бизнес-процессах, а повышение эффективности человеческого труда.

Проекты в области ИИ становятся все разнообразнее и востребованнее, однако инициативы часто тормозятся из-за неудачного выбора архитектуры, недостаточной подготовки и неспособности масштабироваться. Именно поэтому все более актуальной становится роль архитектора в области ИИ, который подготовит и организует запуск

в эксплуатацию решений ИИ. Таким образом актуализируется задача подготовки кадров соответствующей квалификации (архитектор в области искусственного интеллекта) с учетом потребностей бизнеса к квалификации специалистов в области искусственного интеллекта.

Архитекторы в области искусственного интеллекта – это кураторы и владельцы стратегии архитектуры ИИ, связующее звено между дата-сайентистами, дата-инженерами, разработчиками, операционным направлением (DevOps, DataOps, MLOps) и руководителями бизнес-подразделений для управления и масштабирования инициатив в области ИИ. Они тесно сотрудничают с архитекторами предприятий и решений, но в отличие от них, сосредоточены не на широком наборе функций, а на создании надежной корпоративной архитектуры для ИИ.

К ключевым задачам архитектора в области искусственного интеллекта относятся помощь в определении архитектурной стратегии, создание рабочих процессов, определение наборов инструментов и масштабирование операций с ИИ.

Потребность в архитекторах в области ИИ продолжает расти, особенно с переходом бизнеса в онлайн. Образовательная программа в рамках специальности «Архитектор в области искусственного интеллекта», направленная на формирование целостного инженерного видения и практики реализации и применения систем искусственного интеллекта в реальных условиях, с акцентом на технологии программной инженерии обеспечит приобретение следующих профессиональных компетенций:

- способность классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта;
- способность применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез;
- способность участвовать в процессе создания систем искусственного интеллекта на различных этапах жизненного цикла в качестве эксперта и ключевого пользователя;
- способность разрабатывать и применять методы машинного обучения для решения задач;
- способность использовать системы искусственного интеллекта на основе нейросетевых моделей и методов;
- способность использовать инструментальные средства для решения задач машинного обучения;
- способность создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов;

– способность осуществлять сбор и подготовку данных для систем искусственного интеллекта;

– способность создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта.

1.2 Категория слушателей. Требования к уровню подготовки слушателя.

К обучению на программе допускаются: предприниматели, работники и владельцы компаний IT-сектора, имеющие высшее образование или среднее профессиональное (либо получающие высшее или среднее профессиональное образование), а также мотивированные специалисты из других профессиональных сфер и студенты, обучающиеся в области информационных технологий, а также по иным специальностям, которые заинтересованы в получении новых компетенций по специальности «Архитектор в области искусственного интеллекта».

Наличие опыта профессиональной деятельности: без опыта.

Требования к уровню подготовленности, определяемому контрольно-измерительными материалами.

Слушатели должны обладать следующими знаниями, умениями и владеть навыками:

Умение писать программы на каком-либо императивном языке программирования (Java, Python, JavaScript и т.п.).

PYTHON:

Знание синтаксиса языка.

Понимание базовых структур данных.

Иметь базовое представление о работе популярных алгоритмов (бинарный поиск, рекурсия и т. д.).

Базовое представление о скорости работы алгоритмов («O» большое).

Иметь базовое представление о работе с модулем Pandas и NumPy.

SQL:

Знание базового синтаксиса (SELECT, WHERE, GROUP BY, HAVING).

Умение составлять подзапросы и делать все виды JOIN.

Навык работы с вложенными запросами.

ИНФРАСТРУКТУРА:

Умение работать с командной строкой.

Знание базовых команд Linux.

Навыки работы с Git.

Базовое знание Docker.

МАШИННОЕ ОБУЧЕНИЕ:

Знание классических моделей (линейные модели и решающие деревья)

Знание классических метрик качества модели (accuracy, MSE, RMSE).

Требования к компетенциям, которыми должен обладать гражданин при поступлении на программу:

– способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач;

– способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности;

– владеет широкой общей подготовкой (базовыми знаниями) для решения практических задач в области информационных систем и технологий;

– способен использовать современные компьютерные технологии поиска информации для решения поставленной задачи, критического анализа этой информации и обоснования принятых идей и подходов к решению;

– способен выбирать и оценивать способ реализации информационных систем и устройств (программно-, аппаратно- или программно-аппаратно-) для решения поставленной задачи.

1.3. Область профессиональной деятельности.

06 Связь и информационно-коммуникационные технологии (в сферах: анализа, моделирования и формирования интегрального представления стратегий и целей, бизнес-процессов и информационно-технологической инфраструктуры предприятий различной отраслевой принадлежности и различных форм собственности, а также учреждений государственного и муниципального управления; стратегического планирования и управления развитием информационных систем и информационно-

коммуникационных технологий управления предприятием; аналитической поддержки процессов принятия решений для управления предприятием).

1.4 Планируемые результаты обучения.

Программа повышения квалификации разработана с учетом профессиональных стандартов:

«Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты РФ от 6 июля 2020 года №405н;

«Программист», утвержденный приказом Министерства труда и социальной защиты РФ от 20 июля 2022 № 424н.

Программа повышения квалификации разработана с учетом:

ФГОС 38.03.05 Бизнес-информатика, ФГОС 09.04.04. Программная инженерия.

По данной программе приобретаются компетенции универсальной модели компетенций в сфере искусственного интеллекта, разработанной РЭУ им. Г.В. Плеханова в 2021 году в рамках результата Федерального проекта «Искусственный интеллект».

Совершенствуемые и/или формируемые компетенции	Тип компетенции	Планируемые результаты обучения (знать, уметь, владеть — использовать конкретные инструменты)
Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	профессиональная	<p>Знать:</p> <ul style="list-style-type: none"> – определения, историю развития и главные тренды искусственного интеллекта; – архитектуру, принципы, процесс, стадии и методологии разработки решений на основе ИИ; – методы и инструментальные средства решения задач искусственного интеллекта. <p>Уметь:</p> <ul style="list-style-type: none"> – осуществлять поиск информации о новых и перспективных методах анализа больших данных, выполнять сравнительный анализ методов. <p>Владеть:</p> <ul style="list-style-type: none"> – культурой постановки и планирования последовательности

		<p>решения задач анализа данных и классификации;</p> <ul style="list-style-type: none"> – навыками анализа реальных задач из различных предметных областей на уровне отдельных подходов и коллективами алгоритмов
<p>Способен применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез</p>	<p>профессиональная</p>	<p>Знать:</p> <ul style="list-style-type: none"> – методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез; – основы методов оптимизации; – технологии распределенной обработки данных. <p>Уметь:</p> <ul style="list-style-type: none"> – применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез; – строить модель для прогнозирования временного ряда; – использовать цифровые платформы анализа данных для решения профессиональных задач. <p>Владеть:</p> <ul style="list-style-type: none"> – навыками применения алгоритма проблемных ситуаций (алгоритма фаззинга)
<p>Способен участвовать в процессе создания систем искусственного интеллекта на различных этапах жизненного цикла в качестве эксперта и ключевого пользователя</p>	<p>профессиональная</p>	<p>Знать:</p> <ul style="list-style-type: none"> - функции, задачи, навыки, содержание работы архитектора в области искусственного интеллекта; – архитектуру, принципы, процесс, стадии и методологии разработки решений на основе ИИ. <p>Уметь:</p> <ul style="list-style-type: none"> – строить модель для рекомендательных систем; – использовать контейнеризацию и развертывать модель в производство; – решать проблемы переобучения и недообучения алгоритма. <p>Владеть:</p> <ul style="list-style-type: none"> – навыками анализа реальных задач из различных предметных областей на уровне отдельных подходов и коллективами алгоритмов

<p>Способен разрабатывать и применять методы машинного обучения для решения задач</p>	<p>профессиональная</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основные типы, источники больших данных, проблемы безопасности; – методы и программный инструментарий технологий больших данных; – базовые понятия и возможности теории игр для их применения в области машинного обучения и искусственного интеллекта; – системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, SPARK и др.): принципы работы, функционал, язык запросов, преимущества и недостатки; – технологии, методы и инструменты машинного обучения для решения профессиональных задач. <p>Уметь:</p> <ul style="list-style-type: none"> – применять методы машинного обучения для работы с большими данными. <p>Владеть:</p> <ul style="list-style-type: none"> – навыками работы с библиотеками, программными платформами (фреймворками) и программными комплексами машинного обучения
<p>Способен использовать системы искусственного интеллекта на основе нейросетевых моделей и методов</p>	<p>профессиональная</p>	<p>Знать:</p> <ul style="list-style-type: none"> – нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности. <p>Уметь:</p> <ul style="list-style-type: none"> – применять методы машинного обучения и нейронных сетей. <p>Владеть:</p> <ul style="list-style-type: none"> – навыком построения полносвязной нейронной сети для задачи классификации; – основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных
<p>Способен использовать инструментальные средства</p>	<p>профессиональная</p>	<p>Знать:</p> <ul style="list-style-type: none"> – синтаксис и семантику основных

<p>для решения задач машинного обучения</p>		<p>языков искусственного интеллекта (Python, JAVA/C#/C++/Scala) и основные приемы программирования на них;</p> <ul style="list-style-type: none"> – основные алгоритмы, используемые в языках программирования; – принципы, содержание и области применения массово-параллельной обработки и анализа данных для ускорения машинного обучения; – технологии распределенной обработки данных; – принципы и содержание работы с распределенными кластерными системами; – технологии и методы потоковой обработки данных. <p>Уметь:</p> <ul style="list-style-type: none"> – применять системы обработки и анализа больших массивов данных; – применять системы контроля версий Git для разработки решений на основе языков программирования; – применять язык SQL для прикладных решений; – применять язык программирования Python и библиотеки при разработке решений на основе ИИ; – визуализировать анализируемые данные; – использовать шины данных (kafka) при обработке больших данных. <p>Владеть:</p> <ul style="list-style-type: none"> – языками программирования (Python, JAVA/C#/C++/Scala) для реализации методов анализа и структурирования данных; – основными инструментами для работы с данными (Git, Docker, CI/CD, RestAPI, Jupyter notebook, kaggle, Nvidia Cuda, VS, IntelliJIdea); – навыками работы с основными библиотеками Python для визуализации данных (Matplotlib, Seaborn, Plotly); – навыком установки Hadoop на виртуальную машину; – навыком использования Docker; – навыком работы с MapReduce, PySpark;
---	--	--

		<ul style="list-style-type: none"> – навыком работы с NumPy; – математическими методами анализа данных
Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	профессиональная	<p>Знать:</p> <ul style="list-style-type: none"> – архитектуры современных нейронных сетей и их использование для решения профессиональных задач. <p>Уметь:</p> <ul style="list-style-type: none"> – поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов. <p>Владеть:</p> <ul style="list-style-type: none"> – способами улучшить качество модели с помощью методов Feature engineering; – навыком использования библиотеки PyTorch, применения GPU в PyTorch; – навыком построения полносвязной нейронной сети для задачи классификации; – основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных; – методами машинного обучения и нейронных сетей
Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	профессиональная	<p>Знать:</p> <ul style="list-style-type: none"> – методы, инструментальные средства и стандарты статистического и описательного анализа данных; – основные платформы данных (облачные и внутрикорпоративные), применяемые при разработке решений на основе ИИ; – основные виды представления данных: табличные, графовые, временные ряды; – основные уровни представления данных. <p>Уметь:</p> <ul style="list-style-type: none"> – использовать основные статистические методы анализа данных; – проводить EDA — разведочный анализ данных;

		<ul style="list-style-type: none"> – моделировать данные в хранилищах (DWH, Data Lake, NoSQL хранилища данных). <p>Владеть:</p> <ul style="list-style-type: none"> – навыком создания логической и физической базы данных; – навыком BI-аналитики; – языками запросов для обращения к СУБД
Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	профессиональная	<p>Знать:</p> <ul style="list-style-type: none"> – понятие сквозных цифровых субтехнологий искусственного интеллекта; – методы и технологии внедрения сквозных цифровых субтехнологий искусственного интеллекта. <p>Уметь:</p> <ul style="list-style-type: none"> – внедрять технологии компьютерного зрения для обработки изображений и видео, задач биометрии; – внедрять технологии обработки естественного языка для анализа и генерации текстов; – внедрять технологии обучения с подкреплением для задач автоматизации производства и транспорта. <p>Владеть:</p> <ul style="list-style-type: none"> – навыком работы с генеративными моделями GAN для задач генерации; – навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий; – навыком работы с технологией RL — обучения с подкреплением
Способен использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	профессиональная	<p>Знать:</p> <ul style="list-style-type: none"> – понятие сквозных цифровых субтехнологий искусственного интеллекта; – методы и технологии внедрения сквозных цифровых субтехнологий искусственного интеллекта. <p>Уметь:</p> <ul style="list-style-type: none"> – применять технологии компьютерного зрения для обработки изображений и видео, задач биометрии;

		<ul style="list-style-type: none"> – применять технологии обработки естественного языка для анализа и генерации текстов; – применять технологии обучения с подкреплением для задач автоматизации производства и транспорта. <p>Владеть:</p> <ul style="list-style-type: none"> – навыком работы с генеративными моделями GAN для задач генерации; – навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий; – навыком работы с технологией RL — обучения с подкреплением
--	--	---

Критерии для оценки уровня сформированности указанных компетенций и соответствия индикаторам достижения компетенций

Компетенция	Критерии для оценки уровня сформированности компетенций
Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей. Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей проблемной области
Способен применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез	Использует методы автоматизированной генерации и проверки гипотез в сфере исследовательской деятельности. Настраивает, конфигурирует и адаптирует программные средства автоматизированного логического вывода в сфере исследовательской деятельности
Способен участвовать в процессе создания систем искусственного интеллекта на различных этапах жизненного цикла в качестве эксперта и ключевого пользователя	Участвует в коллективной работе по созданию систем искусственного интеллекта в качестве эксперта
Способен разрабатывать и применять методы машинного обучения для решения задач	Проводит анализ требований и определяет необходимые классы задач машинного обучения. Принимает участие в оценке и выборе используемых методов машинного обучения
Способен использовать системы искусственного интеллекта на основе нейросетевых моделей и методов	Осуществляет оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи
Способен использовать	Осуществляет оценку и выбор инструментальных

инструментальные средства для решения задач машинного обучения	средств для решения поставленной задачи
Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	Осуществляет оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи
Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях
Способен использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	Решает прикладные задачи и участвует в реализации проектов в области сквозной цифровой субтехнологии «Компьютерное зрение». Решает прикладные задачи и участвует в реализации проектов в области сквозной цифровой субтехнологии «Обработка естественного языка»
Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	Участвует в реализации проектов в области сквозной цифровой субтехнологии «Компьютерное зрение». Участвует в реализации проектов в области сквозной цифровой субтехнологии «Обработка естественного языка»

2. УЧЕБНЫЙ (ТЕМАТИЧЕСКИЙ) ПЛАН.

№ п/п	Наименование модулей/тем программы	Всего, час	Виды учебных занятий			Формы контроля
			Лекции	Практические занятия	Самостоятельная работа	
Модуль 1. Базовый						
Раздел 1.1 Введение в инженериию больших данных						
1	Тема 1. Введение в профессию. Основные принципы работы искусственного интеллекта и его виды	8	2	4	2	тест, практическая работа
2	Тема 2. Введение в языки программирования. Основные популярные языки для ИИ	8	2	4	2	тест, практическая работа
3	Тема 3. Введение в основные инструменты для работы с данными	8	2	4	2	тест, практическая работа
4	Тема 4. Статистический и описательный анализ данных	8	2	4	2	тест, практическая работа
5	Тема 5. Введение в основные библиотеки Python для визуализации данных. Matplotlib, Seaborn, Plotly	8	2	4	2	тест, практическая работа
6	Тема 6. EDA — разведочный анализ данных. Как найти особенности в данных	10	2	6	2	тест, практическая работа
Раздел 1.2 Основы хранения и обработки данных						
7	Тема 1. Введение в Базы данных и теорию хранения	8	2	4	2	тест, практическая работа

	данных					
8	Тема 2. Язык SQL. Реляционные базы данных, использование SQL	10	2	4	4	тест, практическая работа
9	Тема 3. Введение в Hadoop	8	2	4	2	тест, практическая работа
10	Тема 4. DWH, Data Lake, NoSQL хранилища данных	8	2	4	2	тест, практическая работа
11	Тема 5. Массово-параллельная обработка и анализ данных. Работа с PySpark	8	2	4	2	тест, практическая работа
12	Тема 6. Поточковая обработка данных (data streaming, event processing). Шины данных (kafka)	8	2	4	2	тест, практическая работа
13	Тема 7. Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей	10	2	6	2	тест, практическая работа
14	Тема 8. BI-системы и визуализация данных	8	2	4	2	тест, практическая работа
	Промежуточная аттестация	4		4		Решение кейса
Модуль 2. Профильный						
Раздел 2.1 Архитектура ML						
15	Тема 1. Линейная алгебра, работа с numpy, принцип работы классических моделей ML	8	2	4	2	тест, практическая работа
16	Тема 2. Погружение в ML. Обучение с учителем	10	2	6	2	тест, практическая работа
17	Тема 3. Способы	10	2	6	2	тест,

	улучшить качество модели. Feature engineering. Методы оптимизации					практическая работа
18	Тема 4. Погружение в ML. Обучение без учителя	10	2	6	2	тест, практическая работа
19	Тема 5. Погружение в ML. Временные ряды	10	2	6	2	тест, практическая работа
20	Тема 6. Погружение в ML. Рекомендательные системы	10	2	6	2	тест, практическая работа
21	Тема 7. Развертывание модели в производство. Методы поиска новых производственных режимов (интеллектуальное планирование экспериментов)	8	2	4	2	тест, практическая работа
Раздел 2.2. Архитектура ИИ и нейронных сетей						
22	Тема 1. Введение в нейронные сети	8	2	4	2	тест, практическая работа
23	Тема 2. Сверточные и полносвязные нейронные сети	8	2	4	2	тест, практическая работа
24	Тема 3. CV — компьютерное зрение	10	2	4	4	тест, практическая работа
25	Тема 4. NLP – обработка естественного языка	10	2	4	4	тест, практическая работа
26	Тема 5. GAN — Генеративные модели. Автокодировщики	8	2	4	2	тест, практическая работа
27	Тема 6. RL — обучение с подкреплением	8	2	4	2	тест, практическая работа

28	Тема 7. Процесс, стадии и методологии разработки решений на основе искусственного интеллекта	8	2	4	2	тест, практическая работа
	Промежуточная аттестация	4		4		Решение кейса
	Итоговая аттестация	8		8		Решение кейсов
	Всего часов	260	56	142	62	–

3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК.

Объем программы – 260 часов.

Продолжительность обучения – 3 месяца.

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

Режим занятий: 3-4 часа в день.

Завершение обучения: 25.11.2024

№ п/п	Наименование учебных модулей/ практик/ аттестации	Трудоёмкость (час)	Срок освоения (кол-во учебных дней)
1	Модуль 1. Базовый Раздел 1.1 Введение в инженерию больших данных	50	16
2	Модуль 1. Базовый Раздел 1.2 Основы хранения и обработки данных	68	22
3	Промежуточная аттестация Решение практико-ориентированной задачи (кейса)	4	1
4	Модуль 2. Профильный Раздел 2.1 Архитектура ML	66	22
5	Модуль 2. Профильный Раздел 2.2 Архитектура ИИ и нейронных сетей	60	20
6	Промежуточная аттестация Решение практико-ориентированной задачи (кейса)	4	1
7	Итоговая аттестация	8	2

4. РАБОЧИЕ ПРОГРАММЫ МОДУЛЕЙ УЧЕБНОГО КУРСА ПРОГРАММЫ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ «АРХИТЕКТОР В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА».

Программа повышения квалификации состоит из 2 учебных модулей:

Модуль 1 Базовый.

Раздел 1.1. Введение в инженерии больших данных.

Раздел 1.2. Основы хранения и обработки данных

Модуль 2 Профильный.

Раздел 2.1. Архитектура ML.

Раздел 2.2. Архитектура ИИ и нейронных сетей.

4.1 Рабочая программа модуля 1. Базовый.

Введение в инженерии больших данных. Основы хранения и обработки данных

Цель освоения модуля 1 – приобретение слушателями знаний по основным принципам работы искусственного интеллекта и его видам, изучение основного инструментария для работы с данными; практическое применение методов машинного обучения для работы с большими объемами данных.

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 1:

ПК-1.р. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта

ПК-1.и. Способен применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез

ПК-2.п. Способен участвовать в процессе создания систем искусственного интеллекта, на различных этапах жизненного цикла в качестве эксперта и ключевого пользователя.

ПК-4.р. Способен разрабатывать и применять методы машинного обучения для решения задач.

ПК-7.р. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта.

Планируемые результаты обучения по модулю 1.

По итогам освоения модуля слушатели должны:

Знать:

- определения, историю развития и главные тренды искусственного интеллекта;
- архитектуру, принципы, процесс, стадии и методологии разработки решений на основе ИИ;
- методы и инструментальные средства решения задач искусственного интеллекта;
- методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез;
- основы методов оптимизации;
- технологии распределенной обработки данных;
- функции, задачи, навыки, содержание работы архитектора в области искусственного интеллекта;
- архитектуру, принципы, процесс, стадии и методологии разработки решений на основе ИИ;
- основные типы, источники больших данных, проблемы безопасности;
- методы и программный инструментарий технологий больших данных;
- базовые понятия и возможности теории игр для их применения в области машинного обучения и искусственного интеллекта;
- системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, SPARK и др.): принципы работы, функционал, язык запросов, преимущества и недостатки;
- технологии, методы и инструменты машинного обучения для решения профессиональных задач;
- методы, инструментальные средства и стандарты статистического и описательного анализа данных;
- основные платформы данных (облачные и внутрикорпоративные), применяемые при разработке решений на основе ИИ;
- основные виды представления данных: табличные, графовые, временные ряды;
- основные уровни представления данных.

Уметь:

- осуществлять поиск информации о новых и перспективных методах анализа больших данных, выполнять сравнительный анализ методов;
- применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез;
- строить модель для прогнозирования временного ряда;
- использовать цифровые платформы анализа данных для решения профессиональных задач;

- строить модель для рекомендательных систем;
- использовать контейнеризацию и развертывать модель в производство;
- применять методы машинного обучения для работы с большими данными;
- использовать основные статистические методы анализа данных;
- проводить EDA — разведочный анализ данных;
- моделировать данные в хранилищах (DWH, Data Lake, NoSQL хранилища данных).

Владеть:

- культурой постановки и планирования последовательности решения задач анализа данных и классификации;
- навыками анализа реальных задач из различных предметных областей на уровне отдельных подходов и коллективами алгоритмов;
- навыками применения алгоритма проблемных ситуаций (алгоритма фаззинга);
- навыками работы с библиотеками, программными платформами (фреймворками) и программными комплексами машинного обучения;
- навыком создания логической и физической базы данных;
- навыком BI-аналитики;
- языками запросов для обращения к СУБД.

Учебно-тематический план модуля 1 Базовый.

№ № п/п	Наименование дисциплины, модуля, темы	Трудоемкость		В том числе				Форма контроля
		В зачетных единицах	В часах	Контактная работа			Самостоятельная работа	
				Всего	Лекции	Практические занятия		
1	2	3	4	5	6	7	8	9
Раздел 1.1 Введение в инженерии больших данных								
1	Тема 1. Введение в профессию. Основные принципы работы искусственного интеллекта и его виды		8	6	2	4	2	тест, практическая работа
2	Тема 2. Введение в языки программирования. Основные популярные языки для ИИ		8	6	2	4	2	тест, практическая работа
3	Тема 3. Введение в основные инструменты для работы с данными		8	6	2	4	2	тест, практическая работа
4	Тема 4. Статистический и		8	6	2	4	2	тест,

	описательный анализ данных							практическая работа
5	Тема 5. Введение в основные библиотеки Python для визуализации данных. Matplotlib, Seaborn, Plotly		8	6	2	4	2	тест, практическая работа
6	Тема 6. EDA — разведочный анализ данных. Как найти особенности в данных		10	8	2	6	2	тест, практическая работа
Раздел 1.2 Основы хранения и обработки данных								
7	Тема 1. Введение в Базы данных и теорию хранения данных		8	6	2	4	2	тест, практическая работа
8	Тема 2. Язык SQL. Реляционные базы данных, использование SQL		10	6	2	4	4	тест, практическая работа
9	Тема 3. Введение в Hadoop		8	6	2	4	2	тест, практическая работа
10	Тема 4. DWH, Data Lake, NoSQL хранилища данных		8	6	2	4	2	тест, практическая работа
11	Тема 5. Массово-параллельная обработка и анализ данных. Работа с PySpark		8	6	2	4	2	тест, практическая работа
12	Тема 6. Поточковая обработка данных (data streaming, event processing). Шины данных (kafka)		8	6	2	4	2	тест, практическая работа
13	Тема 7. Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей		10	8	2	6	2	тест, практическая работа
14	Тема 8. BI-системы и визуализация данных		8	6	2	4	2	тест, практическая работа
15	Промежуточная аттестация		4	4		4		Решение кейса
16	Итого по модулю 1		122	92	28	64	30	Решение кейса

Раздел 1.1 Введение в инженерию больших данных.

Тема 1. Введение в профессию. Основные принципы работы искусственного интеллекта и его виды.

Архитектор ИИ: потребность и ценность. Задачи, навыки. Обязанности и функция в команде. Перспективы развития искусственного интеллекта в мире и в РФ. Этические аспекты искусственного интеллекта. Вопросы ответственности и контроля в разработке и использовании искусственного интеллекта.

Тема 2. Введение в языки программирования. Основные популярные языки для ИИ.

Существующие и перспективные методы и программный инструментарий технологий больших данных. Введение в Python. Операторы языка. Типы данных и переменные. Базовые

конструкции и структуры языка. Циклы и условия. Функции. Сетевые запросы. Словари и множества. Библиотеки. Синтаксис языка программирования JAVA/C#/C++/Scala.

Тема 3. Введение в основные инструменты для работы с данными.

Git, Docker, CI/CD, RestAPI. Основные концепции и понятия. Как используется в бизнесе. Платформы: Jupiter notebook, kaggle, Nvidia Cuda. Среды разработки: VS, IntelliJIdea

Тема 4. Статистический и описательный анализ данных.

Введение в математическую статистику и использование ее при анализе данных. Современные методы и инструментальные средства анализа больших данных. Стандарты проведения анализа данных. Как и для чего нужна математическая статистика для анализа данных.

Тема 5. Введение в основные библиотеки Python для визуализации данных. Matplotlib, Seaborn, Plotly.

Построение графиков и визуализация данных с помощью инструментов Matplotlib, Seaborn, Plotly.

Тема 6. EDA — разведочный анализ данных. Как найти особенности в данных.

Планирование и организация аналитических работ с использованием технологий больших данных. Основы планирования аналитических работ. Стандарты проведения анализа данных. Содержание и последовательность выполнения этапов аналитического проекта по исследованию больших данных.

Раздел 1.2 Основы хранения и обработки данных.

Тема 1. Введение в Базы данных и теорию хранения данных.

Типы больших данных: метаданные, полуструктурированные, структурированные, неструктурированные. Виды источников данных: созданные человеком, созданные машинами. Источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования. Основные цели моделирования данных.

Тема 2. Язык SQL. Реляционные базы данных, использование SQL.

Реляционные базы данных для хранения структурированных данных. Сравнение OLAP и OLTP систем. Создание таблиц нормализованных данных. Внедрение денормализованных схем (например, STAR, Snowflake). Использование реляционных баз данных SQL: GreenPlum, Postgres, Oracle. Использование SQL в Big Data.

Тема 3. Введение в Hadoop.

Основы Hadoop. Архитектура Apache Hadoop.

Архитектура HDFS. Структура файлов. Алгоритмы чтения и записи в HDFS. Обзор API для работы с HDFS. Object Storage, сходство и отличия от HDFS.

Тема 4. DWH, Data Lake, NoSQL хранилища данных.

Введение в хранилища данных. Архитектура хранилища данных. Хранилища данных NoSQL, назначение и особенности. Отличия от реляционных баз данных. NoSQL в BigData. Рекомендации по использованию и опыт использования разнородных источников данных и информации в задачах анализа больших данных. Принципы обеспечения безопасности данных в облачных хранилищах данных. Принципы обеспечения безопасности в нереляционных базах данных.

Тема 5. Массово-параллельная обработка и анализ данных Работа с PySpark.

Массово-параллельные базы данных в Big Data. Современная технологическая инфраструктура высокопроизводительных и распределенных вычислений. Технология MapReduce. Работа с распределенной кластерной системой.

Тема 6. Поточковая обработка данных (data streaming, event processing). Шины данных (kafka)

Системы обмена сообщениями Storm и Kafka. Создание конвейеры данных с помощью Storm и Kafka. Шины данных Kafka.

Тема 7. Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей.

Машинное обучение на больших данных. Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация. Методы и модели классификации: логистическая регрессия, деревья решений. Методы и модели регрессии: линейная регрессия, деревья решений. Метрики качества моделей.

Тема 8. BI-системы и визуализация данных.

Что такое BI-системы и зачем нужны. Обзор популярных решений. Введение в Superset. Разработка простого дашборда данных.

Содержание практических занятий по модулю 1.

№ темы	Наименование (содержание) темы, по которой предусмотрено практическое занятие	Формы и методы проведения
Раздел 1.1 Введение в инженерии больших данных		
1	Провести анализ рынка Архитектор в области ИИ: РФ — спарсить данные о вакансиях с сайта hh.ru.	Решение практических заданий
2	Решить практические задачи, используя базовые конструкции и структуры языка.	Решение практических заданий

	Решить практические задачи на Python, используя основные функции для работы со списками и кортежами, структуру словарей и множеств.	
3	Введение в использование Jupyter для чтения и формирования данных на python, как настроить окружение.	Решение практических заданий
4	Математический и числовой анализ с помощью библиотек NumPy, SciPy. Статистический анализ: метод многовариантного тестирования, корреляционный анализ, регрессионный анализ. Статистические методы: параметрические, непараметрические, управляемые, неуправляемые, полууправляемые, кластеризация. Решить практические задачи, используя математическую статистику.	Решение практических заданий
5	Поиск закономерностей с помощью визуализации данных.	Решение практических заданий
6	Использование изученных инструментов для проведения разведочного анализа данных – срезы в данных, визуализация, математическая статистика.	Решение практических заданий
Раздел 1.2 Основы хранения и обработки данных		
1	Необходимо выполнить следующее задание: Создать Логическую и физическую модель базы данных продуктового магазина	Решение практических заданий
2	Установить базу данных PostgreSQL и PGAdmin. Создать базу данных продуктового магазина. Провести ее к 3NF, при необходимости провести ее денормализацию	Решение практических заданий
3	Произвести установку экосистемы Hadoop на виртуальную машину или использовать Docker. Работа с HDFS (интерфейс командной строки)	Решение практических заданий
4	Моделирование данных в Хранилище. Произвести развертывание окружения: СУБД + dbt. Конфигурация проекта dbt, установка модуля dbtVault. Выполнить проектирование детального слоя на базе подхода Data Vault, подготовка метаданных для кодогенерации. Автоматизировать наполнения детального слоя данных с помощью dbt + dbtVault Сформировать витрины данных на основе детального слоя	Решение практических заданий
5	Подключение развернутой БД к PySpark Сформировать витрины данных с помощью PySpark на основе детального слоя	Решение практических заданий
6	Необходимо выполнить следующее	Решение практических

	задание: Установка и настройка Storm и Kafka	заданий
7	Создание собственной модели линейной регрессии, сравнение работы с готовым решением в scikit-learn	Решение практических заданий
8	Необходимо выполнить добавление Superset в DWH. Создать полноценный DWH с Postgres и Superset'ом	Решение практических заданий

Содержание самостоятельной работы слушателей по модулю 1.

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
Раздел 1.1 Введение в инженерии больших данных		
1	Тема 1. Введение в профессию. Основные принципы работы искусственного интеллекта и его виды	Сформулировать задачи архитектора ИИ на основе вакансий. Тест: Функции Архитектора ИИ.
2	Тема 2. Введение в языки программирования. Основные популярные языки для ИИ	Рассмотреть основные алгоритмы, используемые в языках программирования и изучить теорию игр. Тест: Основы синтаксиса и структур в Python.
3	Тема 3. Введение в основные инструменты для работы с данными	Изучить Git, Docker. Тест: Основные инструменты для работы с данными
4	Тема 4. Статистический и описательный анализ данных	Рассмотреть дополнительные способы статистического анализа данных. Тест: Математическая статистика
5	Тема 5. Введение в основные библиотеки Python для визуализации данных. Matplotlib, Seaborn, Plotly	Изучить дополнительные инструменты для построения графиков Тест: Использование графиков при построении, отличие/название
6	Тема 6. EDA — разведочный анализ данных. Как найти особенности в данных	Определить этапы жизненного цикла больших данных Тест: Знание анализа данных
Раздел 1.2 Основы хранения и обработки данных		
1	Тема 1. Введение в Базы данных и теорию хранения данных	Изучить материал: Совместное использование базы данных. Безопасность данных. Тест: Введение в базы данных и теорию хранения данных.

2	Тема 2. Язык SQL. Реляционные базы данных, использование SQL	Изучить материал: Функционал PGAdmin для PostgreSQL. Тест: Реляционные базы данных.
3	Тема 3. Введение в Hadoop	Изучить материал: Распределённые файловые системы. Их составляющие. Их достоинства, недостатки и сфера применения. Тест: Распределенная файловая система HDFS.
4	Тема 4. DWH, Data Lake, NoSQL хранилища данных	Изучить принципы моделирования данных в Data Warehouse Изучить материал: Документно-ориентированная модель данных MongoDB. Тест: Хранилища данных
5	Тема 5. Массово-параллельная обработка и анализ данных. Работа с PySpark	Изучить принципы обеспечения безопасных вычислений в распределенных вычислительных средах Тест: Массово параллельная обработка и анализ данных Работа с PySpark.
6	Тема 6. Поточковая обработка данных (data streaming, event processing). Шины данных (kafka)	Изучить материал: Архитектура обработки данных в реальном времени – Apache Kafka. Тест: Конвейер данных (Storm, Kafka).
7	Тема 7. Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей	Изучить документацию scikit-learn Тест: Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей.
8	Тема 8. BI-системы и визуализация данных	Изучить материал: первые шаги в BI-аналитике. Роль Архитектора данных Тест: BI-системы и визуализация данных

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации.

1. Функции Архитектора ИИ.
2. Основы синтаксиса и структур в Python.
3. Основные инструменты для работы с данными.
4. Математическая статистика.
5. Использование графиков при построении, отличие/название.
6. Знание анализа данных.
7. Введение в Базы данных и теорию хранения данных.
8. Реляционные базы данных.
9. Распределенная файловая система HDFS.

10. Хранилища данных.
11. Массово-параллельная обработка и анализ данных. Работа с PySpark
12. Конвейер данных (Storm, Kafka).
13. Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей.
14. BI-системы и визуализация данных.

Учебно-методическое обеспечение.

Обучающие материалы представлены в виде видеолекций, текстовых и графических материалов, размещенных на образовательной платформе <https://data.1t.ru/>.

Перечень основной и дополнительной учебной литературы.

Нормативно-правовые акты:

1. Национальная программа «Цифровая экономика Российской Федерации», утв. распоряжением Правительства Российской Федерации от 28 июля 2017 г. No 1632-р.
2. Федеральная проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 № 17).

Основная литература:

1. Бураков, М.В. Системы искусственного интеллекта: Учебное пособие / М.В. Бураков. – М.: Проспект, 2019. – 440 с.
2. Карпова, И.П. Базы данных: Учебное пособие / И.П. Карпова. – Санкт-Петербург : Питер, 2021. – 240 с.
3. Мартин Клеппман. Высоконагруженные приложения. Программирование, масштабирование, поддержка / М. Клеппман – СПб : Питер, 2018. – 740 с.
4. Кузнецов, В.А. Системный анализ, оптимизация и принятие решений: Учебник / В. А. Кузнецов, А. А. Черепяхин – Москва : КУРС, 2018. – 256 с.
5. Мышев, А.В. Архитектура для интеллектуальных вычислительных и информационно-измерительных систем с нечеткой средой вычислений // Федеральный институт промышленной собственности, отделение ВПТБ – 2019. – 29 с.
6. Алекс Петров. Распределенные данные. Алгоритмы работы современных систем хранения информации / А. Петров – СПб : Питер, 2021. – 336 с.

7. Дэви Силен, Арно Мейсман, Мохамед Али. Основы Data Science и Big Data. Python и наука о данных / Пер. с англ. Матвеев Е. – СПб : Питер, 2017. – 336 с.: ил – (Серия «Библиотека программиста»). ISBN 978-5-496-02517-1.

Дополнительная литература:

1. Брокман Д. Что мы думаем о машинах, которые думают: Ведущие мировые учёные об искусственном интеллекте / Пер. с англ. М. Исаков – М. : Альпина Пабlishер, 2017. – 550 с.: ил.

2. Бруссард М. Искусственный интеллект: Пределы возможного / Пер. с англ. Е. Арье – М. : Альпина Пабlishер, 2020. – 362 с.

3. Бычков А. Г. Сборник задач по теории вероятностей, математической статистике и методам оптимизации: учебное пособие / А.Г. Бычков. — Москва : Форум: ИНФРА-М, 2022. — 192 с. — (Среднее профессиональное образование).

4. Душкин Р.В. Искусственный интеллект / Р.В. Душкин – М. : ДМК Пресс, 2022. – 380 с.: ил. – ISBN:978-5-97060-787-9.

5. Фальк Ким. Рекомендательные системы на практике Практическое пособие / Пер. с англ. Д. М. Павлов. – Москва : ДМК Пресс, 2020. – 448 с.

6. О’Коннелл М. Искусственный интеллект и будущее человечества / Пер. с англ М. Кудряшова – М. : Бомбора, 2019. – 280 с.: ил.

7. Мартин Ф. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей / Пер. с англ. И. Рузмайкина. – СПб : Питер, 2019. – 400 с.: ил.

8. Нархид Ния, Шапира Гвен, Палино Тодд. Apache Kafka. Поточковая обработка и анализ данных – СПб : Питер, 2021. – 320 с.: ил.

9. Поляков В. М. Методы оптимизации: учебное пособие / Поляков В. М., Агаларов З. С. — Москва : Дашков и К, 2022. — 86 с. — ISBN 978-5-394-05003-9.

10. Рязанов С. И. Искусственный интеллект как множество - классификация искусственных интеллектов // Вузовская наука в современных условиях: сборник материалов 54-й Научно-технической конференции: в 3-х ч. Ульяновск : Ульяновский государственный технический университет, 2020. Ч. 1. С. 72-75.

Учебно-методические и информационные материалы:

1. Асанов, В.Л. Стратегическое управление территориальным развитием — архитектурный менеджмент, администрирование: монография / В. Л. Асанов. — Москва : Издательство «Юрайт», 2023. — 275 с.

2. Большие данные в социальных и гуманитарных науках: Сб. обзоров и рефератов / РАН. ИНИОН. Центр науч.-информ. исслед. по науке, образованию и

технологиям; отв. ред. – Гребенщикова Е.Г. – М. , 2019. – 193 с. – (Сер.: Наука, образование и технологии). ISBN 978-5-248-00912-1.

3. Загорулько, Ю. А. Искусственный интеллект. Инженерия знаний: учебное пособие для вузов / Ю. А. Загорулько, Г. Б. Загорулько. — Москва : Издательство Юрайт, 2022. — 93 с. — (Высшее образование). — ISBN 978-5-534-07198-6.

4. Нагородская, В.Б. Новые технологии (блокчейн / искусственный интеллект) на службе права: научнометодическое пособие / под ред. Л. А. Новоселовой. — Москва : Проспект, 2019. — 128 с.

5. Цифровая экономика от теории к практике: как российский бизнес использует искусственный интеллект [Электронный ресурс] / Исслед. РАЭК / НИУ ВШЭ при поддержке Microsoft. – 2019. – 66 с. – Режим доступа: <http://raec.ru/upload/files/190715-ii.pdf> (Дата обращения: 27.03.2024).

6. Бизнес переходит на искусственный интеллект. / РБК+. Решения. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс] – Режим доступа: <https://plus.rbc.ru/news/638ce98f7a8aa9f3126daaa2> (Дата обращения: 27.03.2024).

7. Расставить нейросети. / РБК+. Инновации. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс] – Режим доступа: <https://plus.rbc.ru/news/638ce67f7a8aa9e27b22f26e> (Дата обращения: 27.03.2024).

8. Технологии позволяют учитывать специфику каждой отрасли / РБК+. От первого лица. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс] – Режим доступа: <https://plus.rbc.ru/news/638ced6c7a8aa9e28f7bf148> (Дата обращения: 27.03.2024).

Информационное сопровождение.

Электронные образовательные ресурсы:

1. Воронцов, К. В. Машинное обучение: курс лекций // MachineLearning.ru. - URL:

[http://www.recognition.su/wiki/index.php?title=Машинное_обучение_\(курс_лекций%2C_К.В.Воронцов\)](http://www.recognition.su/wiki/index.php?title=Машинное_обучение_(курс_лекций%2C_К.В.Воронцов)) (Дата обращения: 27.03.2024).

2. Проектирование баз данных: распределенные базы и хранилища данных. Агрегирование // Национальный открытый университет «ИНТУИТ». – URL: http://www.intuit.ru/studies/professional_retraining/953/courses/214/lecture/5508/ (Дата обращения: 27.03.2024).

3. Бесплатные материалы по Data Engineering от преподавателей МФТИ // МФТИ. – [Электронный ресурс] – Режим доступа: <https://fpmi-edu.ru/free-de> (Дата обращения: 27.03.2024).
4. Курс «Big Data и Data Science: начни погружение с нуля» // Stepik. – [Электронный ресурс] – Режим доступа: <https://stepik.org/course/101687/promo> (Дата обращения: 27.03.2024).
5. Open Machine Learning Course // mlcourse.ai. – [Электронный ресурс] – Режим доступа: <https://mlcourse.ai/book/index.html> (Дата обращения: 27.03.2024).
6. NLP-курс от ШАД // Lena Voita. – [Электронный ресурс] – Режим доступа: https://lena-voita.github.io/nlp_course.html (Дата обращения: 27.03.2024).
7. Бесплатный курс по Deep Learning от МФТИ // ФПМИ МФТИ. – [Электронный ресурс] – Режим доступа: <https://dls.samcs.ru> (Дата обращения: 27.03.2024).
8. Обучение Python и его библиотекам // W3Schools Spaces – [Электронный ресурс] – Режим доступа: <https://www.w3schools.com/python/default.asp> (Дата обращения: 27.03.2024).

Электронные информационные ресурсы.

1. Курсы по искусственному интеллекту // УНТИ 2035– [Электронный ресурс] – Режим доступа: <https://ai.2035.university/> – (Дата обращения: 27.03.2024).
2. Национальный проект «Цифровая экономика» // Национальные проекты – [Электронный ресурс]. – Режим доступа: <https://национальныепроекты.рф/projects/tsifrovaya-ekonomika> (Дата обращения: 27.03.2024)
3. Федеральный проект «Искусственный интеллект»// Национальный проект «Цифровая экономика» // Национальные проекты – [Электронный ресурс]. – Режим доступа: <https://национальныепроекты.рф/projects/tsifrovaya-ekonomika/p-iskusstvennyu-intellekt-p> (Дата обращения: 27.03.2024)
4. Курсы от образовательной платформы ООО «1Т» – [Электронный ресурс] – Режим доступа: <https://data.1t.ru/> (Дата обращения: 27.03.2024).
5. Искусственный интеллект // РБК – [Электронный ресурс]. – Режим доступа: <https://trends.rbc.ru/trends/tag/ai> (Дата обращения: 27.03.2024)

Описание системы оценки качества освоения дисциплины.

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль предусмотрен в ходе изучения каждой темы.

Формами текущего контроля являются тесты и выполнение практических работ.

Тесты содержат не менее 5 вопросов с одним или несколькими правильными ответами. За каждый правильный ответ ставится 1 балл. Критерием прохождения теста является получение не менее 75% правильных ответов.

Примеры тестов.

Модуль 1. Базовый.

Раздел 1.2 Основы хранения и обработки данных.

Тема: Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей.

Вопросы:

1. Какой вид задач решает алгоритм машинного обучения, называемый «классификация»?
 - a) Прогнозирование числового значения
 - b) Выделение групп похожих объектов
 - c) Определение принадлежности объекта к определенному классу**
 - d) Разбиение выборки на обучающую и тестовую
2. Какой из этих методов не относится к классическим моделям машинного обучения?
 - a) Линейная регрессия
 - b) Решающее дерево
 - c) Нейронные сети**
 - d) Логистическая регрессия
3. Какая метрика используется для оценки качества моделей классификации?
 - a) RMSE
 - b) R^2
 - c) F1-score**
 - d) MSE
4. Какая из этих задач является задачей регрессии?
 - a) Определение наличия определенного объекта на изображении
 - b) Кластеризация пользователей по просмотру фильмов на онлайн-платформе
 - c) Прогнозирование цены на недвижимость**
 - d) Определение тональности отзывов на товары
5. Какой метод машинного обучения решает задачу поиска похожих объектов?
 - a) Линейная регрессия
 - b) Решающие деревья

с) Метод опорных векторов

d) К-ближайших соседей

Выполнение практических работ оценивается в бинарной системе: зачтено / не зачтено.

Зачтено: задача решена, могут быть недочеты и неточности в решении.

Не зачтено: задача не решена.

Примеры практических работ.

Модуль 1. Базовый.

Раздел 1.2 Основы хранения и обработки данных.

Тема: 1. Введение в Базы данных и теорию хранения данных.

Задача: Создание логической и физической модели базы данных.

Спроектировать БД для выбранной предметной области согласно примеру, рассмотренному на практике.

1. Описать предметную область
2. Выделить ключевые объекты системы
3. Провести инфологическое проектирование

Ссылку на репозиторий с кодом прислать в чат с преподавателем.

Промежуточная аттестация проводится в формате решения практико-ориентированной задачи (кейса).

Примеры практико-ориентированных задач (кейсов).

Модуль 1.

Задание 1.

В качестве исходных данных используется датасет: **test.csv** и **train.csv**.

1. Загрузить данные, обработать ошибки в данных – заполнить либо удалить пропуски, удалить дубликаты.

2. Провести EDA-анализ данных. Построить графики – выбрать оптимальные для числовых и категориальных признаков. Найти возможные закономерности в данных. Провести статистический анализ данных – сравнить выборки, где целевая переменная 1 и где 0.

3. Произвести предобработку данных – выполнить масштабирование и кодирование данных для подготовки к обучению моделей машинного обучения. Вся предобработка должна осуществляться и на train и на test выборках.

4. Обучить модель логистической регрессии.
5. Обучить модель дерева решений.
6. Обучить модель решающих деревьев.
7. Сравнить качества моделей и выбрать наиболее лучшую модель для предсказания.

Цель: достичь метрики точности (ассигасу) более 0,7.

Задание 2.

Спроектируйте схему данных и постройте аналитическую витрину:

1. Развернуть базу данных и загрузить данные. Загрузить только 70% данных.
2. Подключить к разработанной БД любой BI-инструмент.
3. Вывести основные характеристики данных в витрину.
4. Разработать модель машинного обучения (линейную регрессию) для предсказания значений.
5. Вывести метрику качества модели в витрину.
6. Добавить новые данные, выделенные ранее, и рассмотреть новые метрики. Вывести разницу между новыми метриками и старыми.

Ссылку на репозиторий с кодом прислать в чат с преподавателем.

Критерии и шкала оценивания:

0–4 балла: имеются содержательные и логические ошибки, решение кейса не найдено;
5–6 баллов: решение кейса в целом найдено, но оно неоптимально и/или имеются логические ошибки;

7–8 баллов: решение кейса найдено, но имеются неточности в решении;

9–10 баллов: решение кейса найдено, ошибки отсутствуют.

Максимально возможное число баллов за работу – 10:

не менее 9 баллов – «отлично»;

7–8 баллов – «хорошо»;

5–6 баллов – «удовлетворительно»;

0–4 балла – «неудовлетворительно».

4.2 Рабочая программа модуля 2. Профильный.

Архитектура ML. Архитектура ИИ и нейронных сетей.

Цель освоения модуля 2 – приобретение слушателями знаний и практического опыта по созданию и поддержке системы искусственного интеллекта на основе нейросетевых моделей и методов, овладение языками программирования (Python, JAVA/C#/C++/Scala) и другими инструментами для реализации методов анализа и структурирования больших данных.

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 2:

ПК-5.п. Способен использовать системы искусственного интеллекта на основе нейросетевых моделей и методов.

ПК-5.р. Способен использовать инструментальные средства для решения задач машинного обучения.

ПК-6.р. Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов.

ПК-9.р. Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта.

ПК-8.п. Способен использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта.

Планируемые результаты обучения по модулю 2.

По итогам освоения модуля слушатели должны:

Знать:

- нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности;
- принципы, содержание и области применения массово-параллельной обработки и анализа данных для ускорения машинного обучения;
- технологии распределенной обработки данных;
- принципы и содержание работы с распределенными кластерными системами;
- технологии и методы потоковой обработки данных;
- архитектуры современных нейронных сетей и их использование для решения профессиональных задач;
- понятие сквозных цифровых субтехнологий искусственного интеллекта;
- методы и технологии внедрения сквозных цифровых субтехнологий искусственного интеллекта;

- понятие сквозных цифровых субтехнологий искусственного интеллекта;
- методы и технологии внедрения сквозных цифровых субтехнологий искусственного интеллекта.

Уметь:

- применять методы машинного обучения и нейронных сетей;
- применять системы обработки и анализа больших массивов данных;
- использовать шины данных (kafka) при обработке больших данных;
- поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов;
 - внедрять технологии компьютерного зрения для обработки изображений и видео, задач биометрии;
 - внедрять технологии обработки естественного языка для анализа и генерации текстов;
 - внедрять технологии обучения с подкреплением для задач автоматизации производства и транспорта;
 - применять технологии компьютерного зрения для обработки изображений и видео, задач биометрии;
 - применять технологии обработки естественного языка для анализа и генерации текстов;
 - применять технологии обучения с подкреплением для задач автоматизации производства и транспорта.

Владеть:

- навыком построения полносвязной нейронной сети для задачи классификации;
- основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных;
- способами улучшить качество модели с помощью методов Feature engineering;
- навыком использования библиотеки PyTorch, применения GPU в PyTorch;
- навыком построения полносвязной нейронной сети для задачи классификации;
- основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных;
- методами машинного обучения и нейронных сетей;
- навыком работы с генеративными моделями GAN для задач генерации;

- навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий;
- навыком работы с технологией RL — обучения с подкреплением;
- навыком работы с генеративными моделями GAN для задач генерации;
- навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий;
- навыком работы с технологией RL — обучения с подкреплением.

Учебно-тематический план модуля 2 Профильный.

№ п/п	Наименование дисциплины, модуля, темы			В том числе				Самостоятельная работа	Форма контроля
		В зачетных единицах	В часах	Всего	Контактная работа				
					Лекции	Практические занятия			
1	2	3	4	5	6	7	8	9	
Раздел 2.1 Архитектура ML									
1	Тема 1. Линейная алгебра, работа с NumPy, принцип работы классических моделей ML		8	6	2	4	2	тест, практическая работа	
2	Тема 2. Погружение в ML. Обучение с учителем		10	8	2	6	2	тест, практическая работа	
3	Тема 3. Способы улучшить качество модели. Feature engineering. Методы оптимизации		10	8	2	6	2	тест, практическая работа	
4	Тема 4. Погружение в ML. Обучение без учителя		10	8	2	6	2	тест, практическая работа	
5	Тема 5. Погружение в ML. Временные ряды		10	8	2	6	2	тест, практическая работа	
6	Тема 6. Погружение в ML. Рекомендательные системы		10	8	2	6	2	тест, практическая работа	
7	Тема 7. Развертывание модели в производство. Методы поиска новых производственных режимов (интеллектуальное планирование экспериментов)		8	6	2	4	2	тест, практическая работа	

Раздел 2.2 Архитектура ИИ и нейронных сетей								
8	Тема 1. Введение в нейронные сети		8	6	2	4	2	тест, практическая работа
9	Тема 2. Сверточные и полносвязные нейронные сети		8	6	2	4	2	тест, практическая работа
10	Тема 3. CV — компьютерное зрение		10	6	2	4	4	тест, практическая работа
11	Тема 4. NLP – обработка естественного языка		10	6	2	4	4	тест, практическая работа
12	Тема 5. GAN — Генеративные модели. Автокодировщики		8	6	2	4	2	тест, практическая работа
13	Тема 6. RL — обучение с подкреплением		8	6	2	4	2	тест, практическая работа
14	Тема 7. Процесс, стадии и методологии разработки решений на основе искусственного интеллекта		8	6	2	4	2	тест, практическая работа
15	Промежуточная аттестация		4	4		4		Решение кейса
16	Итого по модулю 2		130	98	28	70	32	Решение кейсов

Раздел 2.1 Архитектура ML.

Тема 1. Линейная алгебра, работа с NumPy, принцип работы классических моделей ML.

Для чего нужна линейная алгебра в ML. Матрицы и работа матриц в классических моделях. Обработка матриц с помощью библиотеки NumPy. Математическое моделирование. Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма.

Тема 2. Погружение в ML. Обучение с учителем.

Обучение с учителем. Задача классификации. Метод опорных векторов. Решающие деревья. Случайный лес. K-ближайших соседей (KNN). Бустинги.

Принципы работы таких моделей. Кросс-валидация.

Тема 3. Способы улучшить качество модели. Feature engineering. Методы оптимизации.

Добавление дополнительных параметров в данные, извлечение признаков, парсинг из внешних источников. Борьба с переобучением. Подбор гиперпараметров модели. Автоматический поиск параметров с помощью поиска по сетке. Поиск ассоциативных правил. Нормализация данных. Кросс-валидация – дополнительные способы валидации. Бэггинг и стэкинг.

Тема 4. Погружение в ML. Обучение без учителя.

Обучение без учителя. Метод наименьших квадратов. Задача кластеризации данных. Поиск аномалий.

Тема 5. Погружение в ML. Временные ряды.

Прогнозирование временных рядов — основные методы, способы валидации, метрики качества моделей.

Тема 6. Погружение в ML. Рекомендательные системы.

Рекомендательные системы — основные методы, способы валидации, метрики. Информационный поиск.

Тема 7. Развертывание модели в производство. Методы поиска новых производственных режимов (интеллектуальное планирование экспериментов).

Основные проблемы и важность данных задач со стороны бизнеса. Все шаги создания сервиса и интеграции ML-модели в эксплуатацию. Методы управления жизненным циклом информационно-технологической инфраструктуры организации. Методы управления проектами создания информационно-технологической инфраструктуры организации. Основы компьютерных сетевых технологий.

Раздел 2.2 Архитектура ИИ и нейронных сетей.

Тема 1. Введение в нейронные сети.

Введение в нейронные сети. Алгоритм обратного распространения ошибки для обучения нейронных сетей. Основные архитектуры и задачи нейронных сетей. Производная, оптимизация.

Тема 2. Сверточные и полносвязные нейронные сети.

Продвинутое обучение нейронных сетей. Многослойный перцептрон. Архитектура CNN, свёрточные нейронные сети. Разница между моделями и как устроена архитектура.

Тема 3. CV — компьютерное зрение.

Классификация, сегментация изображений. Распознавание объектов на изображении. Распознавание текста на изображении. Способы оптимизации. Анализ изображений и видео с помощью методов искусственного интеллекта.

Тема 4. NLP – обработка естественного языка.

Основные задачи и для чего используется. Что такое векторное представление слов, мешок слов и N-граммы, TF-IDF, эмбединги. Виды нейронных сетей для задач NLP. Использование BERT для построения эмбединга слов. Анализ естественного языка с помощью методов искусственного интеллекта.

Тема 5. GAN — Генеративные модели. Автокодировщики.

Введение в генерацию текстов. Текстовая генерация с трансформеров. Автоэнкодеры – автокодировщики и использование их для снижения размерности и поиска закономерностей в данных.

Тема 6. RL — обучение с подкреплением.

Reinforcement Learning – обучение с подкреплением, кто такой агент. Основные библиотеки.

Тема 7. Процесс, стадии и методологии разработки решений на основе искусственного интеллекта.

Разработка и согласование технического задания на создание методической и технологической инфраструктуры больших данных. Управление архитектурой организации. Методы управления жизненным циклом информационно-технологической инфраструктуры организации. Методы управления проектами создания информационно-технологической инфраструктуры организации. Инструменты и методы согласования с заказчиками требований к результатам аналитических исследований с использованием технологий больших данных. Архитектура и принципы работы промышленных решений, созданных на основе ИИ. Области применения искусственного интеллекта.

Содержание практических занятий по модулю 2.

№ темы	Наименование (содержание) темы, по которой предусмотрено занятие семинарского типа	Формы и методы проведения
Раздел 2.1 Архитектура ML		
1	Предобработка данных и построение нескольких классических моделей и выбор лучшей модели на данных	Решение практических заданий
2	Предобработка данных и построение нескольких классических моделей и выбор лучшей модели на данных	Решение практических заданий
3	Построить новые признаки в данных, обучить новые модели и подобрать гиперпараметры модели	Решение практических заданий
4	Поиск аномалий в данных, сегментация PCA Уменьшение размерности данных	Решение практических заданий
5	Построить модель для прогнозирования	Решение практических

	временного ряда	заданий
6	Построить рекомендательную систему	Решение практических заданий
7	Построение полного жизненного цикла модели на основе производственного примера	Решение практических заданий
Раздел 2.2 Архитектура ИИ и нейронных сетей		
1	Построение собственной небольшой модели линейной регрессии на основе библиотеки PyTorch	Решение практических заданий
2	Построение полносвязной нейронной сети для задачи классификации на основе датасета MNIST	Решение практических заданий
3	Построение сверточной нейронной сети для задачи многоклассовой классификации	Решение практических заданий
4	Анализ естественного языка с помощью методов искусственного интеллекта. Воспользоваться TF-IDF или эмбедингами для анализа токсичности текстов	Решение практических заданий
5	Построение собственной модели генерации текста с оригинальным стилем и создание чат-бота	Решение практических заданий
6	Создание собственной модели RL на основе игры Space Invaders Ссылка: Space Invaders	Решение практических заданий
7	Провести классификацию знаний. Изучить заданную предметную область и построить модель знаний в виде графа	Решение практических заданий

Содержание самостоятельной работы слушателей по модулю 2.

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
Раздел 2.1 Архитектура ML		
1	Тема 1. Линейная алгебра, работа с NumPy, принцип работы классических моделей ML	Изучить материал: Генетические алгоритмы, эволюционное программирование Тест: Метрики качества моделей
2	Тема 2. Погружение в ML. Обучение с учителем	Изучить материал: Принцип работы Наивного Байесовского классификатора Тест: Обучение с учителем
3	Тема 3. Способы улучшить качество модели. Feature engineering. Методы оптимизации	Изучить материал: Автоматический поиск параметров с помощью рандомизированного поиска

		Библиотеки Optuna Тест: Методы оптимизации.
4	Тема 4. Погружение в ML. Обучение без учителя	Изучить материал: <ul style="list-style-type: none"> • Восстановление регрессии данных • Иерархическая кластеризация • Модельная кластеризация • Сеточная кластеризация Тест: Кластеризация данных
5	Тема 5. Погружение в ML. Временные ряды	Изучить материал: ARIMA и SARIMAX для прогнозирования рядов Тест: Временные ряды
6	Тема 6. Погружение в ML. Рекомендательные системы	Изучить материал: Дополнительные библиотеки для построения рекомендательной системы Тест: Рекомендательные системы
7	Тема 7. Развертывание модели в производство. Методы поиска новых производственных режимов (интеллектуальное планирование экспериментов)	Изучить материал: Основы обеспечения информационной безопасности Тест: Принятие бизнес решений на основе данных
Раздел 2.2 Архитектура ИИ и нейронных сетей		
1	Тема 1. Введение в нейронные сети	Изучить материал: Использование GPU в PyTorch Библиотека TensorFlow и основные различия Тест: Нейронные сети
2	Тема 2. Сверточные и полносвязные нейронные сети	Изучить материал: Основные популярные архитектуры сверточных сетей Тест: Сверточные сети
3	Тема 3. CV — компьютерное зрение	Изучить тему: Использование CV в видео. Тест: Computer vision - основы
4	Тема 4. NLP – обработка естественного языка	Изучить материал: Модели трансформеров и что такое self-attention Тест: Основные понятия NLP
5	Тема 5. GAN — Генеративные модели. Автокодировщики	Изучить материал: Использование автокодировщиков для снижения размерности и построения визуализации данных Тест: Основы состязательно-генеративных моделей
6	Тема 6. RL — обучение с подкреплением.	Изучить материал: Как используется RL в бизнесе. Как обучаются автопилоты на основе RL Тест: основные концепции RL
7	Тема 7. Процесс, стадии и методологии разработки решений на основе искусственного интеллекта	Изучить материал: Регламенты по оформлению требований к результатам аналитических исследований с использованием технологий больших данных

		Технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии Современный опыт использования анализа больших данных Тест: Роль искусственного интеллекта в бизнесе.
--	--	---

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации.

1. Метрики качества моделей.
2. Обучение с учителем.
3. Методы оптимизации.
4. Кластеризация данных.
5. Временные ряды.
6. Рекомендательные системы.
7. Принятие бизнес-решений на основе данных.
8. Нейронные сети.
9. Сверточные сети.
10. Computer vision – основы.
11. Основные понятия NLP.
12. Основы состязательно-генеративных моделей.
13. Основные концепции RL.
14. Роль искусственного интеллекта в бизнесе.

Учебно-методическое обеспечение.

Обучающие материалы представлены в виде видеолекций, текстовых и графических материалов, размещенных на образовательной платформе <https://data.1t.ru/>.

Перечень основной и дополнительной учебной литературы.

Нормативно-правовые акты:

1. Национальная программа «Цифровая экономика Российской Федерации», утв. распоряжением Правительства Российской Федерации от 28 июля 2017 г. No 1632-р.
2. Федеральный проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 № 17).

Основная литература:

1. Бураков, М.В. Системы искусственного интеллекта: Учебное пособие / М.В. Бураков. – М. : Проспект, 2019. – 440 с.
 2. Мартин Ф. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей / Пер. с англ. И. Рузмайкина. – СПб : Питер, 2019. – 400 с.: ил.
 3. Конвински Энди, Венделл Патрик, Захария Матей, Карау Холден. Изучаем Spark. Молниеносный анализ данных – Москва : ДМК Пресс, 2015. – 304 с.
 4. Мышев, А.В. Архитектура для интеллектуальных вычислительных и информационно-измерительных систем с нечеткой средой вычислений – 2019.
 5. Алекс Петров. Распределенные данные. Алгоритмы работы современных систем хранения информации – СПб : Питер, 2021. – 336 с.
- Дополнительная литература:
1. Болотова, Ю.А. Методы и алгоритмы интеллектуальной обработки цифровых изображений: учебное пособие / Ю. А. Болотова, А. А. Друки, В. Г. Спицын; Национальный исследовательский Томский политехнический университет (ТПУ). — Томск : Изд-во ТПУ, 2016. — 207 с.: ил. — Библиогр.: с. 206. — ISBN 978-5-4387-0710-3.
 2. Брокман Д. Что мы думаем о машинах, которые думают: Ведущие мировые учёные об искусственном интеллекте / Пер. с англ. М. Исаков – М. : Альпина Паблишер, 2017. – 550 с.: ил.
 3. Бруссард М. Искусственный интеллект: Пределы возможного / Пер. с англ. Е. Арье – М. : Альпина Паблишер, 2020. – 362 с.
 4. Брэдшоу Шеннон, Брэзил Йон, Ходоров Кристина. MongoDB: полное руководство. Мощное и масштабируемая система управления базами данных. 3-я редакция / Пер. с англ. Д. А. Беликова. – М. : ДМК Пресс, 2020. - 540 с.: ил.
 5. Томас Дэвенпорт. Внедрение искусственного интеллекта в бизнес-практику. Преимущества и сложности / Пер. с англ. З. Мамедьяров. – М. : Альпина Паблишер, 2021. – 316 с.
 6. Душкин Р.В. Искусственный интеллект / Р.В. Душкин – М. : ДМК Пресс, 2022. – 380 с.: ил. – ISBN:978-5-97060-787-9
 7. Фальк Ким. Рекомендательные системы на практике Практическое пособие / Пер. с англ. Д. М. Павлов. – Москва : ДМК Пресс, 2020. – 448 с.
 8. О’Коннелл М. Искусственный интеллект и будущее человечества / Пер. с англ М. Кудряшова – М. : Бомбора, 2019. – 280 с.: ил.

9. Маркус Г. Искусственный интеллект: Перезагрузка. Как создать машинный разум, которому действительно можно доверять / Г. Маркус, Э. Дэвис. – М. : Альпина Паблшер, 2021. – 328 с.
 10. Мартин Ф. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей / Пер. с англ. И. Ружмайкина. – СПб. : Питер, 2019. – 400 с.: ил.
 11. Нархид Ния, Шапира Гвен, Палино Тодд. Apache Kafka. Поточковая обработка и анализ данных – СПб. : Питер, 2021. – 320 с.
 12. Пиковер К. Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей / Пер. с англ. А. Ефримова. – М. : Синдбад, 2022. – 250 с.
 13. Цзэн Мин. Как Alibaba использует искусственный интеллект в бизнесе: Сетевое взаимодействие и анализ данных / Пер. с кит. В. Ионов. – М. : Альпина Паблшер, 2022. – 360 с.: ил.
 14. Поляков В. М. Методы оптимизации: учебное пособие / Поляков В. М., Агаларов З. С. — Москва : Дашков и К, 2022. — 86 с. — ISBN 978-5-394-05003-9.
 15. Рязанов С. И. Искусственный интеллект как множество - классификация искусственных интеллектов // Вузовская наука в современных условиях: сборник материалов 54-й Научно-технической конференции: в 3-х ч. – Ульяновск : Ульяновский государственный технический университет, 2020. Ч. 1. – 72-75 с.
 16. Сергеев, Л. И. Цифровая экономика: учебник для вузов / Л. И. Сергеев, Д. Л. Сергеев, А. Л. Юданова ; под редакцией Л. И. Сергеева. — 2-е изд., перераб. и доп. — М. : Издательство Юрайт, 2023. — 437 с. — (Высшее образование). — ISBN 978-5-534-15797-0.
- Учебно-методические и информационные материалы:
1. Асанов, В. Л. Стратегическое управление территориальным развитием — архитектурный менеджмент, администрирование: монография / В. Л. Асанов. — Москва : Издательство Юрайт, 2023. — 275 с. — (Актуальные монографии). — ISBN 978-5-534-12772-0.
 2. Большие данные в социальных и гуманитарных науках: Сб. обзоров и рефератов / РАН. ИНИОН. Центр науч.-информ. исслед. по науке, образованию и технологиям; отв. ред. – Гребенщикова Е.Г. – М., 2019. – 193 с. – (Сер.: Наука, образование и технологии). ISBN 978-5-248-00912-1.
 3. Загорулько, Ю. А. Искусственный интеллект. Инженерия знаний: учебное пособие для вузов / Ю. А. Загорулько, Г. Б. Загорулько. — Москва : Издательство Юрайт, 2022. — 93 с. — (Высшее образование). — ISBN 978-5-534-07198-6.

4. Нагородская, В.Б. Новые технологии (блокчейн / искусственный интеллект) на службе права: научнометодическое пособие / под ред. Л. А. Новоселовой. — Москва : Проспект, 2019. — 128 с.

5. Цифровая экономика от теории к практике: как российский бизнес использует искусственный интеллект [Электронный ресурс] / Исслед. РАЭК / НИУ ВШЭ при поддержке Microsoft. — 2019. — 66 с. — Режим доступа: <http://raec.ru/upload/files/190715-ii.pdf> (Дата обращения: 27.03.2024).

6. Бизнес переходит на искусственный интеллект. / РБК+. Решения. #1 Искусственный интеллект, 5 декабря 2022. — [Электронный ресурс] — Режим доступа: <https://plus.rbc.ru/news/638ce98f7a8aa9f3126daaa2> (Дата обращения: 27.03.2024).

7. Расставить нейросети. / РБК+. Инновации. #1 Искусственный интеллект, 5 декабря 2022. — [Электронный ресурс] — Режим доступа: <https://plus.rbc.ru/news/638ce67f7a8aa9e27b22f26e> (Дата обращения: 27.03.2024).

8. Технологии позволяют учитывать специфику каждой отрасли / РБК+. От первого лица. #1 Искусственный интеллект, 5 декабря 2022. — [Электронный ресурс] — Режим доступа: <https://plus.rbc.ru/news/638ced6c7a8aa9e28f7bf148> (Дата обращения: 27.03.2024).

Статьи:

1. Алейникова, Ю.В.; Матвеев, В.В. Цифровая экосистема. Анализ применения искусственного интеллекта // Здоровье – основа человеческого потенциала: проблемы и пути их решения. — 2020. — Т. 15. — № 3. — С. 1480-1487.

2. Безгачев, Ф.В. Особенности обеспечения кибербезопасности посредством технологий искусственного интеллекта / В сборнике: Актуальные проблемы борьбы с преступностью: вопросы теории и практики. Материалы XXV международной научно-практической конференции, в 2-х частях. — Красноярск, 2022. — С. 43-44.

3. Городнова, Н.В. Применение искусственного интеллекта в бизнес-сфере: современное состояние и перспективы // Вопросы инновационной экономики. — 2021. — Том 11. — № 4. — С. 1473-1492. — doi: 10.18334/vines.11.4.112249.

4. Дин Но., Афанасьев Г.И. Состояние и перспективы применения искусственного интеллекта в визуализирующей диагностике заболеваний легких // E-Scio. — 2022. — № 4 (67). — С. 653-664.

5. Доржиева, В.В. Цифровизация промышленности: роль искусственного интеллекта и возможности для России // Вопросы инновационной экономики. — 2022. — Т. 12. № 4. — С. 2383-2394.

6. Коженков, А.О.; Середя, А.М.; Поливаев, О.И. Искусственный интеллект в современном автомобилестроении / В сборнике: Теория и практика инновационных технологий в АПК. Материалы национальной научно-практической конференции. – Воронеж, 2022. – С. 210-216.
7. Красов, А.В.; Штеренберг, С.И.; Фахрутдинов, Р.М.; Рыжаков, Д.В.; Пестов, И.Е. Анализ информационной безопасности предприятия на основе сбора данных пользователей с открытых ресурсов и мониторинга информационных ресурсов с использованием машинного обучения [Электронный ресурс] // Т-Comm. 2018. №10. – Режим доступа: <https://cyberleninka.ru/article/n/analiz-informatsionnoy-bezopasnosti-predpriyatiya-na-osnove-sbora-dannyh-polzovateley-s-otkrytyh-resursov-i-monitoringa> (Дата обращения: 27.03.2024).
8. Куницкая, О.М. Правовое регулирование искусственного интеллекта применительно к задачам развития экономики / В сборнике: Право и инновации: новые вызовы технологической революции. Материалы II Приволжского юридического конгресса. – Уфа, 2022. – С. 77-91.
9. Лазарева, М.М.; Калюкарин, А.В. Преимущества искусственного интеллекта. Применение в банковской сфере // Вопросы устойчивого развития общества. – 2022. – № 9. – С. 312-316.
10. Лазько, Н.В. Применение искусственного интеллекта в нефтегазовой отрасли // Мехатроника, автоматика и робототехника. – 2023. – № 11. – С. 155-158.
11. Львович, И.Я. Проблемы методологии проектирования интеллектуальных информационных систем // Информационные технологии в управлении, автоматизации и мехатронике. – 2020. – С. 120-123.
12. Машошин, А.И. Применение искусственного интеллекта при создании систем управления силами ВМФ // Морская радиоэлектроника – 2022. – № 2 (80). – С. 22-25.
13. Нарушения при работе с информационными системами Россельхознадзора будет выявлять искусственный интеллект // Пищевая промышленность. – 2022. – № 3. – С. 76.
14. Пиляй, И.В. Применение систем искусственного интеллекта для оценки времени и стоимости строительного проекта // Строительство и архитектура. – 2023. – Т. 11. – № 1. – С. 19.
15. Репин, И. С. Перспективные направления использования искусственного интеллекта в оперативно-розыскной деятельности // Журнал правовых и экономических исследований. – 2022. – № 2. – С. 125-130.

16. Субботин, А.В.; Тагилова, Л.Ф. Математическое моделирование информационных процессов проектирования интеллектуальных систем на основе использования метода Мамдани // Информационные технологии как основа прогрессивных научных исследований. – 2020. – С. 95-99.

17. Шиллер, А.В. Место этической системы в архитектуре искусственного интеллекта // Вестник Томского государственного университета. – 2020. – №. 456. – С. 99-103.

Информационное сопровождение.

Электронные образовательные ресурсы:

1. Воронцов, К. В. Машинное обучение: курс лекций // MachineLearning.ru. – [Электронный ресурс] – Режим доступа: [http://www.recognition.ru/wiki/index.php?title=Машинное_обучение_\(курс_лекций%2C_К.В.Воронцов\)](http://www.recognition.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций%2C_К.В.Воронцов)) (Дата обращения: 27.03.2024).

2. Проектирование баз данных: распределенные базы и хранилища данных. Агрегирование // Национальный открытый университет «ИНТУИТ». – [Электронный ресурс] – Режим доступа: http://www.intuit.ru/studies/professional_retraining/953/courses/214/lecture/5508/ (Дата обращения: 27.03.2024).

3. Бесплатные материалы по Data Engineering от преподавателей МФТИ // МФТИ. – [Электронный ресурс] – Режим доступа: <https://fpmi-edu.ru/free-de> (Дата обращения: 27.03.2024).

4. Курс «Big Data и Data Science: начни погружение с нуля» // Stepik. – [Электронный ресурс] – Режим доступа: <https://stepik.org/course/101687/promo> (Дата обращения: 27.03.2024).

5. Open Machine Learning Course / mlcourse.ai. – [Электронный ресурс]. – Режим доступа: <https://mlcourse.ai/book/index.html> (Дата обращения: 27.03.2024).

6. NLP-курс от ШАД // Lena Voita. – [Электронный ресурс] – Режим доступа: https://lena-voita.github.io/nlp_course.html (Дата обращения: 27.03.2024).

7. Бесплатный курс по Deep Learning от МФТИ // ФПМИ МФТИ. – [Электронный ресурс] – Режим доступа: <https://dls.samcs.ru> (Дата обращения: 27.03.2024).

8. Обучение Python и его библиотекам // W3Schools Spaces – [Электронный ресурс] – Режим доступа: <https://www.w3schools.com/python/default.asp> (Дата обращения: 27.03.2024).

Электронные информационные ресурсы:

1. Курсы по искусственному интеллекту // УНТИ 2035– [Электронный ресурс]. – Режим доступа: <https://ai.2035.university/> – (Дата обращения: 27.03.2024).
2. Национальный проект «Цифровая экономика» // Национальные проекты – [Электронный ресурс]. – Режим доступа: <https://национальныепроекты.рф/projects/tsifrovaya-ekonomika> (Дата обращения: 27.03.2024)
3. Федеральный проект «Искусственный интеллект»// Национальный проект «Цифровая экономика» // Национальные проекты – [Электронный ресурс]. – Режим доступа: <https://национальныепроекты.рф/projects/tsifrovaya-ekonomika/p-iskusstvennyu-intellekt-p> (Дата обращения: 27.03.2024).
4. Курсы от образовательной платформы ООО «1Т» – [Электронный ресурс]. – Режим доступа: <https://data.1t.ru/> (Дата обращения: 27.03.2024).
5. Искусственный интеллект // РБК – [Электронный ресурс]. – Режим доступа: <https://trends.rbc.ru/trends/tag/ai> (Дата обращения: 27.03.2024).

Описание системы оценки качества освоения модуля 2.

Контроль результатов освоения модуля осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль предусмотрен в ходе изучения каждой темы.

Формами текущего контроля являются тесты и выполнение практических работ.

Тесты содержат не менее 5 вопросов с одним или несколькими правильными ответами. За каждый правильный ответ ставится 1 балл. Критерием прохождения теста является получение не менее 75% правильных ответов.

Примеры тестов.

Модуль 2. Профильный.

Раздел 2.2 Архитектура ИИ и нейронных сетей.

Тема: Введение в нейронные сети.

Вопросы:

1. Что такое нейронная сеть?
 - а) Механизм для ускорения вычислений
 - б) Система управления роботами
 - в) Модель машинного обучения, основанная на принципах работы мозга**
 - д) Способ кодирования информации
2. Какие основные элементы составляют нейронную сеть?

a) Входной слой, скрытый слой, выходной слой

b) Сенсоры, активаторы, рецепторы

c) Компьютер, процессор, память

d) Клетки, ткани, органы

3. Что такое функция активации?

a) Метод оптимизации в процессе обучения

b) Функция, используемая для оценки точности ответа

c) Функция, используемая для расчета ошибки

d) Функция, используемая для определения выхода нейрона

4. Какие типы задач можно решать с помощью нейронных сетей?

a) Классификация, регрессия, кластеризация

b) Перевод языков, генерация текста, распознавание речи

c) Анализ изображений, обработка звука, работа с текстом

d) Все вышеперечисленные

5. Какие алгоритмы обучения используются в нейронных сетях?

a) Метод градиентного спуска, стохастический градиентный спуск

b) Метод опорных векторов, метод главных компонент

c) Методы машинного обучения, случайный поиск

d) Решающие деревья, метод k-ближайших соседей

Выполнение практических работ оценивается в бинарной системе: зачтено / не зачтено.

Зачтено: задача решена, могут быть недочеты и неточности в решении.

Не зачтено: задача не решена.

Примеры практических работ.

Модуль 2. Профильный.

Раздел 2.2 Архитектура ИИ и нейронных сетей.

Тема: 4. NLP – обработка естественного языка.

Задача:

Использовать датасет twitter текстов для определения токсичности текстов.

Предобработать данные – разделить данные, очистить.

Векторизировать текст.

Выбрать любую модель машинного обучения и обучить на данных.

Рассмотреть метрики качества, вывести результат.

Ссылку на репозиторий с кодом прислать в чат с преподавателем.

Промежуточная аттестация проводится в формате решения практико-ориентированной задачи (кейса).

Примеры практико-ориентированных задач (кейсов).

Модуль 2.

Задание 1.

В качестве исходных данных используется датасет: `test_prices-regression.csv` и `train_prices-regression.csv`.

Загрузить данные, обработать ошибки в данных – заполнить либо удалить пропуски, удалить дубликаты.

Провести EDA – анализ данных. Построить графики – выбрать оптимальные для числовых и категориальных признаков. Найти закономерности изменения цен на дома и квартиры. Построить таблицы корреляции – по Пирсону и Phik-корреляции. Изучить все графики и таблицы на наличие закономерностей.

Произвести анализ данных с помощью метода понижения размерности.

Произвести предобработку данных – выполнить масштабирование и кодирование данных для подготовки к обучению моделей машинного обучения. Вся предобработка должна осуществляться и на `train` и на `test` выборках.

Обучить модель линейной регрессии. Изучить самые важные признаки (`feature importances`) – на основе этих данных построить новые признаки и сравнить качество модели.

Обучить модель градиентного бустинга и подобрать лучшие гиперпараметры с помощью `GridSearchCV`.

Сравнить качества моделей и выбрать наиболее лучшую модель для предсказания.

Для оценки валидности модели – проверить модель на адекватность с помощью `dummy-regressor`. Если результат выбранной итоговой модели будет хуже `dummy` – выяснить причину и произвести все шаги заново с 6 по 7.

Цель: достичь метрики `RMSE` менее 2000.

Задание 2.

Кластеризация новостей.

Выгрузить (спарсить) данные с любого сайта новостей – за последнюю неделю.

Если есть краткое описание новости – использовать его, если нет – использовать полный текст новости.

Векторизировать текст с помощью любых библиотек Python.

С помощью векторизованного текста – кластеризовать текст на подгруппы.
Выделить основные слова по кластеру.

Сравнить результат кластеризации с фактическим разбиением по темам на сайте.

Произвести аналитику новостей.

Ссылку на репозиторий с кодом прислать в чат с преподавателем.

Критерии и шкала оценивания:

0–4 балла: имеются содержательные и логические ошибки, решение кейса не найдено;

5–6 баллов: решение кейса в целом найдено, но оно неоптимально и/или имеются логические ошибки;

7–8 баллов: решение кейса найдено, но имеются неточности в решении;

9–10 баллов: решение кейса найдено, ошибки отсутствуют.

Максимально возможное число баллов за работу – 10:

не менее 9 баллов – «отлично»;

7–8 баллов – «хорошо»;

5–6 баллов – «удовлетворительно»;

0–4 балла – «неудовлетворительно».

5. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ.

Специфика организационных действий и педагогических условий.

Для достижения планируемых результатов обучение строится с использованием следующих:

методов: case-study, метод проектов, модульное обучение, проблемное обучение, контекстное обучение;

форм: лекции с использованием мультимедиа, практические занятия, самостоятельная работа.

Кроме того, обучение строится с применением **технологий** электронного обучения и дистанционных образовательных технологий.

Кадровое обеспечение программы (преподавательский состав).

К реализации программы привлечены представители образовательных организаций высшего образования и представители компаний со стажем работы в области искусственного интеллекта и в смежных областях.

Представители образовательных организаций высшего образования имеют высшее образование, ученую степень кандидата или доктора наук, стаж научно-педагогической работы более трех лет, а также не выполняют функции иностранного агента.

Представители компаний со стажем работы в области искусственного интеллекта и в смежных областях имеют опыт решения практических задач с использованием технологий искусственного интеллекта более 3 лет в течение последних 10 лет в профильной компании или в профильном подразделении, а также не выполняют функции иностранного агента.

Члены преподавательского состава имеют за последние 3 года научные публикации, соответствующие направлению данной программы, в журналах, включенных в перечень ВАК (К1 и К2), а также в журналах, включенных в «Белый список» Минобрнауки РФ 1-го квартиля.

№ п/п	Фамилия, имя, отчество (при наличии)	Место основной работы и должность, ученая степень и ученое звание (при наличии)	Ссылки на веб-страницы с портфолио (при наличии)
1	Борисов Вадим Владимирович <i>«Белый список» – 4 шт., из них 1-го квартиля – 1шт.</i>	Профессор кафедры вычислительной техники филиала НИУ «МЭИ» в г. Смоленске, д.т.н., профессор	

	<i>ВАК К1 – 2 шт.</i>		
2	Бен Режеб Тауфик Бен Камель <i>«Белый список» – 1 шт.</i> <i>ВАК К1 – 5 шт.</i> <i>ВАК К2 – 1 шт.</i>	Декан факультета «Радио и телевидение» МТУСИ	
3	Дроздов Сергей Николаевич	Доцент Института компьютерных технологий и информационной безопасности ФГАОУ ВО «Южный федеральный университет», к.т.н.	
4	Захарова Алена Александровна <i>«Белый список» – 4 шт., из них 1-го кварття – 1 шт.</i> <i>ВАК К1 – 1 шт.</i> <i>ВАК К2 – 1 шт.</i>	Главный научный сотрудник ИПУ РАН, д.т.н., доцент	
5	Ерохин Кирилл Сергеевич	Руководитель отдела Data Science, ООО «1Т»	https://data.1t.ru/erohin
6	Ерохин Виталий Александрович <i>ВАК К1 – 1 шт.</i>	Ведущий специалист по ИИ, ООО «1Т»	https://data.1t.ru/erohin_v
7	Кузин Антон Алексеевич	Старший преподаватель, Департамент математического и компьютерного моделирования, Институт математики и компьютерных технологий МТУСИ	https://1t.ru/teacher-kuzin
8	Бондаренко Евгений Владимирович <i>ВАК К1 – 1 шт.</i>	Ведущий специалист по ИИ, ООО «1Т»	
9	Козлова Анна Сергеевна	Специалист по ИИ, ООО «1Т»	
10	Юров Юрий Аркадьевич	Специалист по ИИ, ООО «1Т»	
11	Дон Юрий Никитич	Разработчик машинного обучения, ООО «Рубиус тех»	

12	Софьина Каролина Германовна <i>ВАК К1 – 1 шт.</i>	Специалист по ИИ, ООО «IT»	
----	---	----------------------------	--

Материально-технические условия реализации программы

Вид занятий	Наименование оборудования, программного обеспечения
Лекционные занятия	Персональный компьютер с установленным на нем: Windows 10-11, x64/x86; от 8 Gb RAM; от 128 Gb SSD/HDD, монитор от 15"; сетевой интерфейс Fast Ethernet 100 Мбит; веб-браузеры Google Chrome, Mozilla Firefox, Opera, Microsoft Edge, Яндекс.Браузер и др.
Практические занятия, самостоятельная работа, промежуточная и итоговая аттестация	Персональный компьютер с установленным на нем: Windows 10 и выше, x64/x86; от 8 Gb RAM; от 128 Gb SSD/HDD, монитор от 15"; сетевой интерфейс Fast Ethernet 100 Мбит; веб-браузеры Google Chrome, Mozilla Firefox, Opera, Microsoft Edge, Яндекс.Браузер и др. Anaconda 2.7 или 3.5 Доступ к облачным вычислительным ресурсам.

Материально-технические условия соответствуют действующим санитарным и противопожарным правилам и нормам.

При проведении учебных занятий с применением дистанционных образовательных технологий (ДОТ) у слушателя должен быть персональный компьютер, оснащенный аудиоколонками, с доступом в сеть интернет и установленным видеоплеером, способным воспроизводить видеофайлы.

Выдаваемый документ при успешном освоении программы.

Удостоверение о повышении квалификации ООО «IT»

6. СИСТЕМА ОЦЕНКИ КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ.

В систему оценки качества освоения программы входят:

- 1) текущий контроль;
- 2) промежуточная аттестация;
- 3) итоговая аттестация.

Формы, методы проведения и оценочные материалы текущего контроля и промежуточной аттестации представлены в соответствующих рабочих программах модулей.

Для зачисления на курс потенциальному слушателю необходимо пройти входную диагностику (вступительное испытание).

5.1 Входная диагностика (вступительное испытание)

Входная диагностика (вступительное испытание) проводится в тестовой форме.

1. Какой командой можно вывести количество строк в файле txt? Файл называется «text.txt»

- a. mkdir text.txt
- b. cat text.txt | wc -l**
- c. cat text.txt
- d. rm text.txt | wc -l

2. Как просмотреть содержимое папки через командную строку?

- a. ld
- b. cd
- c. ws
- d. ls**

3. Какая команда выведет директорию, в которой сейчас находится пользователь?

- a. pd
- b. pwd**
- c. wd
- d. cd

4. Через какую команду осуществляется переход из каталога в каталог?

- a. pd

b. rmdir

c. cd

d. wd

5. С помощью команды `Ls -L /bin` получили в начале записи `-rwxr-xr-x (755)`. Что это значит?

a. Владелец имеет доступ на чтение, запись и выполнение; группа и другие могут только читать и выполнять.

b. Владелец, группа, прочие пользователи имеют все права на файл

c. Владелец файла может читать, изменять и выполнять файл

d. Только группа владельца имеет право изменять файл

6. Если объем выборки достаточно велик (больше 100), то такая выборка является репрезентативной?

a. Да

b. Зависит от способа формирования выборки

c. Нет

d. 100 – небольшая выборка

7. Признак данных имеет такие значения – 0.01, 1, 2, 0.0003. Какой это тип данных?

a. Количественный

b. Булевый

c. Категориальный

d. Порядковый

8. Что такое дисперсия в понятии анализа данных?

a. Стандартное отклонение

b. Сумма квадратов разностей значений и среднего, деленное на количество значений

c. Обозначение одного из типа данных

d. Что-то связанное с погодой

9. Что такое математическое ожидание случайной величины?

a. Сумма всех значений случайной величины, умноженных на их вероятности

- b. случайная величина
- c. вероятность признака
- d. вероятностное пространство

10. С какой вероятностью произвольно взятая буква из русского алфавита окажется гласной?

- a. 1/33
- b. 10/33**
- c. 23/33
- d. 33/33

11. Что такое Data Science?

- a. Изучение цифровых устройств.
- b. Анализ данных с использованием математических и статистических методов.**
- c. Наука о поведении людей в цифровом мире.
- d. Программирование игр для компьютеров.

12. Что такое регрессия в анализе данных?

- a. Метод определения тенденции развития событий.
- b. Метод моделирования измеряемых данных и исследования их свойств**
- c. Метод расчета вероятности события.
- d. Метод защиты данных от несанкционированного доступа.

13. Какими категориями делятся алгоритмы обучения с учителем?

- a. Кластеризация и классификация.
- b. Регрессия и классификация.**
- c. Кластеризация и регрессия.
- d. Нет правильного ответа.

14. Что такое "очистка данных"?

- a. Очистка записей от пыли и мусора.
- b. Преобразование неструктурированных данных в удобочитаемый формат.
- c. Обработка и устранение ошибок, неточностей и пропусков в данных.**
- d. Поиск скрытых связей между данными.

15. Для чего используют метод кластеризации?

- a. **Для разделения объектов на группы на основе их характеристик.**
- b. Для предсказания цены на товары в будущем.
- c. Для поиска выбросов в наборе данных.
- d. Нет правильного ответа.

16. Что такое SVM в Data Science?

- a. Стандартная векторная модель.
- b. **Метод опорных векторов.**
- c. Метод стохастического градиентного спуска.
- d. Алгоритм случайного леса.

17. Что такое переобучение?

- a. **Эффект, при котором модель слишком хорошо запоминает обучающие данные и плохо работает на новых данных.**
- b. Метод преподавания в школах.
- c. Способ уменьшения размерности данных.
- d. Нет правильного ответа.

18. Что такое F-мера в машинном обучении?

- a. Мера точности алгоритма классификации.
- b. Мера полноты алгоритма классификации.
- c. **Среднее гармоническое между точностью и полнотой.**
- d. Нет правильного ответа.

19. Что такое кросс-валидация?

- a. Метод оптимизации параметров модели
- b. Статистический метод для анализа зависимости переменных
- c. Разделение данных на обучающую и контрольную выборку
- d. **Метод проверки качества модели на нескольких фолдах (валидациях)**

20. Что делает функция describe() в pandas.DataFrame?

- a. **Рассчитывает описательные статистики для каждого признака**

- b. Сортирует данные по заданному признаку
- c. Удаляет строки или столбцы с пропущенными значениями
- d. Создает новый столбец на основе имеющегося

Пороговое значение для успешного прохождения вступительного испытания (не менее 65% от общего количества результатов выполнения заданий).

Итоговая аттестация.

Итоговая аттестация проходит в форме решения практико-ориентированных задач (кейсов).

Практико-ориентированные задачи (кейсы) для итоговой аттестации предоставлены организациями, работающими в области искусственного интеллекта и смежных областях (ООО «Альмира»), а также организациями, применяющими технологии интеллектуальной обработки данных для решения своих производственных задач (ООО «Верконт Сервис», ООО СП «Содружество», ООО ЭЦ «Социология и аналитика», ПК «ИТ Союз»).

Практико-ориентированная задача (кейс) 1. Анализ характеристик и свойств виноградного сока.

Описание задачи:

Даны табличные данные о свойствах красного вина.

Задача – разработать алгоритм прогнозирования качества вина на основе характеристик.

Этапы задачи:

1. Провести очистку и предобработку данных.
2. Провести анализ данных – исследовательский и статистический анализ данных, рассмотреть распределение качества вина по характеристикам.
3. Выбрать алгоритм машинного обучения – классификатор или регрессия.
4. Обучить модель и выбрать наилучшую.
5. Выделить самые важные признаки при формировании модели.
6. Проверить выбранную модель на адекватность (проверка со среднего значения предсказания).

В качестве результата работы программного продукта необходимо оформить тетрадку jupyter notebook и выложить решение на github.

Практико-ориентированная задача (кейс) 2. Разработка модели классификации цветов.

Описание задачи:

Даны картинки и папки – в них находятся картинки с описанием цветов и их сорта.

Задача – спроектировать полный цикл разработки модели нейронной сети и ее эксплуатации.

Этапы задачи:

- 1 Спроектировать БД для хранения такого типа данных.
 - 2 Провести анализ данных – рассмотреть дисбаланс классов.
 - 3 Разработать сверточную нейронную сеть для классификации изображений цветов – добиться наилучшей метрики accuracy.
 - 4 Обучить модель и сохранить, проверить качество модели на новых данных.
 - 5 Разработать сервис для автоматической классификации изображений.
- В качестве результата работы программного продукта необходимо разработать контейнер docker и выложить результат на github.

Практико-ориентированная задача (кейс) 3. Разработка модели автоматической фильтрации негативных сообщений.

Описание задачи:

Дан датасет с негативными и положительными комментариями.

Задача – разработать модель и выложить рабочую модель в облачный сервис для быстрого доступа.

Этапы задачи:

1. Спроектировать БД для хранения такого типа данных.
2. Провести анализ данных – рассмотреть, какие чаще всего слова встречаются, о чем чаще говорят, какие тематики сообщений.
3. Автоматизировать анализ данных текстов и выводить результат в виде dashboard.
4. Предобработать данные и обучить модель
5. Разработать pipeline – полный цикл работы модели.
6. Разработать контейнер docker с обученной моделью и внедрить решение на любой облачный сервис.

В качестве результата работы программного продукта выложить результат на github и предоставить доступ к фильтру. Необходимый результат – на вход подаются примеры текстов, выход – результат классификатора.

Практико-ориентированная задача (кейс) 4. Разработка модели генерации поздравлений.

Описание задачи:

Нужно обучить генеративную модель для генерации текстов поздравлений.

Задача – разработать модель и выложить рабочую модель.

Этапы задачи:

1 Спроектировать БД для хранения запросов и результатов запросов.

2 Обучить генеративную модель.

3 Автоматизировать и максимально ускорить работу модели.

4 Разработать dashboard с результатами запросов модели – внутренней отладке и аналитике работы модели.

5 Разработать чат-бота, где пользователь отправляет «затравку» — о чем поздравление, а результат – полный текст поздравления.

В качестве результата работы программного продукта выложить результат на github и предоставить инструкцию.

Практико-ориентированная задача (кейс) 5. Разработка модели сбора новостей из источников и их анализа.

Описание задачи:

Задача – нужно обучить модель для полного анализа цикла новостей за любой промежуток времени.

Этапы задачи:

1 Спроектировать БД для хранения новостей по их источникам, а также для хранения результатов работы модели.

2 Спарсить все новости за месяц из выбранных источников.

3 Провести полный анализ текстов – кластеризировать тексты, выделить основные темы из результатов работы модели машинного обучения. Сохранить результаты в БД.

4 Автоматизировать работу сбора данных и работы модели через Airflow.

5 Сформировать витрину данных.

6 Разработать docker-контейнер с полным циклом работы модели.

В качестве результата работы программного продукта выложить результат на github и предоставить инструкцию.

Практико-ориентированная задача (кейс) 6. Разработка модели кредитного скоринга.

Описание задачи: обучить модель для статистической оценки вероятности мошеннических действий со стороны потенциального заемщика.

Этапы задачи:

1 Провести EDA предоставленных данных по заемщикам.

2 Собрать полный отчет в Jupyter notebook – основные закономерности в данных и статистический анализ. Сформировать витрину данных.

3 Предобработать данные – очистить от дубликатов/пропусков или заполнить пропуски.

4 Обучить модель машинного обучения на основе данных.

5 Насытить данные для лучшей работы модели.

6 Улучшить качество модели с помощью подбора гиперпараметров.

7 Сохранит модель для будущей работы.

8 Разработать контейнер с автоматической обработкой данных и обучением модели.

В качестве результата работы программного продукта выложить результат на github и предоставить инструкцию.

Практико-ориентированная задача (кейс) 7. Разработка модели прогнозирования спроса.

Цель: разработать модель прогнозирования спроса.

Для разработки модели, которая будет прогнозировать спрос на определенный товар в будущем, необходимо выполнить следующие задачи.

Этапы задачи:

1. **Сбор данных.** Необходимо собрать данные о продажах товара за определенный период времени (например, год). Для этого можно использовать любой публичный источник данных. Важно, чтобы данные были представлены в удобном формате и содержали достаточное количество информации о продажах товара (например, количество проданных единиц товара в каждый день/неделю/месяц и цена за единицу товара).

2. **Подготовка данных.** Данные, собранные на предыдущем этапе, нужно подготовить для дальнейшего анализа. Необходимо провести первичную обработку данных, оценить их качество и, при необходимости, произвести очистку данных от выбросов, ошибок и пропусков.

3. **Анализ данных.** После подготовки данных необходимо провести исследовательский анализ данных, чтобы понять характеристики и зависимости между переменными. На этом этапе можно использовать статистические методы анализа данных, построить графики для лучшего понимания данных.

4. **Моделирование.** На этом этапе необходимо построить модель, которая будет прогнозировать спрос на товар в будущем. Для этого можно использовать различные методы машинного обучения. Важно выбрать модель, которая будет давать наилучшие результаты прогнозирования.

5. **Развитие и тестирование модели.** Необходимо провести тестирование модели на новых данных, чтобы убедиться в ее качестве и готовности к использованию в реальных условиях.

6. **Автоматизировать работу сбора данных и работы модели.** Для автоматизации работы с данными и моделью можно использовать инструмент Airflow, который позволит автоматически запускать задачи и процессы, связанные с разработкой модели.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

3. Примеры данных, которые использовались для разработки модели, и результаты ее работы на этой выборке данных.

Практико-ориентированная задача (кейс) 8. Разработка модели генерации картинок.

Цель: разработать модель генерации картинок.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1 **Выбрать готовую модель.** Взять готовую модель с сайта huggingface с любым стилем генерации изображений.

2 **Тестирование модели.** Разверните модель локально и протестируйте скорость генерации. Подберите такие размеры картинки, чтобы генерация не была долгой.

3 **Моделирование.** Дообучите модель и создайте новый промт на основе желаемого результата.

4 **Развитие и тестирование модели.** После нужно использовать генератор для генерации новых изображений. Нужно поэкспериментировать с размером шума и другими параметрами, чтобы получить нужные изображения. Готовую модель необходимо сохранить и разработать небольшой GUI (любая готовая библиотека, можно gradio).

5 **Разработать контейнер Docker.** Разработать контейнер с готовой моделью и интерфейсом.

Результат работы программного продукта

1 Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2 Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 9. Разработка модели чат-бота для автоматического заказа еды в ресторане.

Проект: Разработка чат-бота для автоматического заказа еды в ресторане.

Этапы разработки:

1. Исследование. Определение наиболее подходящих инструментов и библиотек для создания чат-бота.

2. Сбор и обработка данных. Определение списка ресторанов, которые будут подключены к чат-боту. Сбор информации об их меню, ценах и доступных способах оплаты. Обработка данных для последующего использования в чат-боте.

3. Проектирование архитектуры чат-бота. Определение структуры диалогового интерфейса чат-бота. Проектирование архитектуры базы данных для хранения информации о заказах. Разработка алгоритмов обработки сообщений пользователя и генерации ответов от чат-бота.

4. Разработка и тестирование. Разработка чат-бота с использованием выбранных инструментов и библиотек. Внесение корректировок в чат-бот на основе результатов тестирования.

5. Развертывание и запуск в производство. Деплой чат-бота на сервер

Результаты работы — документация по проекту, которая включает в себя:

1. Описание технологии и библиотек, используемых в проекте.
2. Структура базы данных и ее описание.
3. Описание архитектуры чат-бота и принципы его работы.
4. Описание процесса развертывания и запуска.
5. Инструкции для пользователей чат-бота.

Практико-ориентированная задача (кейс) 10. Распознавание рукописного ввода: создайте модель машинного обучения, которая может распознавать символы рукописного ввода и преобразовывать их в текст.

Цель: разработать модель распознавания рукописного ввода OCR.

Для разработки модели, необходимо выполнить следующие задачи.

Этапы задачи:

1. **Сбор данных.** Необходимо собрать различные изображения рукописного ввода с любых источников.

2. **Хранение данных.** Разработать архитектуру Data lake для хранения изображений. Можно использовать любого вендора.

3. **Разработка модели.** Разработать модель распознавания текста – собрать необходимую структуру нейронной сети и обучить.

4. **Оценка и тестирование.** Протестировать работу модели на новых изображениях.

5. **Разработать контейнер Docker.** Разработать контейнер с готовой моделью.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 11. Рекомендательная система: создайте модель, которая может рекомендовать пользователю товары или услуги на основе его предыдущих покупок или интересов.

Цель: разработать рекомендательной модели для подбора фильмов.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. **Сбор данных.** Необходимо собрать различные данные по фильмам.

2. **Хранение данных.** Разработать архитектуру базы данных для хранения информации о фильмах.

3. **Обработка данных.** Проанализировать данные и обработать их – векторизировать текст.

4. **Разработка модели.** Разработать модель рекомендательной системы и подобрать необходимые гиперпараметры модели.

5. **Оценка и тестирование.** Протестировать работу модели.

6. **Разработать контейнер Docker.** Разработать контейнер с готовой моделью.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 12. Обнаружение аномалий: создайте модель, которая может определять аномалии или выбросы в данных, которые могли бы указывать на потенциальные проблемы в бизнес-процессах.

Описание задачи: обучить модель для оценки вероятности аномальных действий со стороны потенциального плательщика.

Этапы задачи:

1. Провести EDA предоставленных данных по плательщикам.
2. Собрать полный отчет в Jupyter notebook – основные закономерности в данных и статистический анализ. Сформировать витрину данных.
3. Обучить модель для поиска аномалий временного ряда плательщика.
4. Сохранить модель с лучшими характеристиками.
5. Разработать контейнер с автоматической обработкой данных и обучением модели.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.
2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 13. Идентификация лиц: создайте модель, которая может распознавать лица на фотографиях или в режиме реального времени, что может быть полезным для систем безопасности или учета рабочего времени.

Цель: обучение нейронной сети для распознавания лиц в организации.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. **Хранение данных.** Разработать архитектуру базы данных для хранения информации о людях и биометрии лиц.
2. **Обработка данных.** Обработать данные для задачи биометрии – векторизовать изображения в тензоры.
3. **Разработка модели.** Разработать нейронную сеть и провалидировать модель.

4. Развитие и тестирование модели. Необходимо провести тестирование модели на новых данных, оптимизировать скорость работы модели для работы с видео.

5. Разработка GUI. Разработать или собрать готовый интерфейс, который будет сохранять изображения из видео и идентифицировать людей.

6. Разработать контейнер Docker. Разработать контейнер с готовой моделью.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 14. Добавление эффектов визуализации: создайте модель, которая может добавлять эффекты визуализации фотографии.

Цель: разработать модель генерации картинок.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. Разработка модели. Разработка модели нейронной сети, которая будет принимать на вход два изображения – картинку и картинку с любым стилем. На выходе первая картинка будет в таком же стиле, как вторая.

2. Развитие и тестирование модели. Нужно поэкспериментировать с размером шума и другими параметрами, чтобы получить нужные стили изображения. Готовую модель необходимо сохранить и разработать небольшой GUI (любая готовая библиотека, можно gradio).

3. Разработать контейнер Docker. Разработать контейнер с готовой моделью и интерфейсом.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 15. Генерация музыки: создайте модель, которая может генерировать музыку на основе заданных параметров, например, стиля или настроения.

Цель: разработать модель генерации музыки.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. **Разработать модель.** Взять готовую модель или разработать модель генерации музыки.

2. **Тестирование модели.** Разверните модель локально и протестируйте скорость генерации.

3. **Моделирование.** Дообучите модель и создайте новый промпт на основе желаемого результата.

4. **Развитие и тестирование модели.** После нужно использовать генератор для генерации новых изображений. Нужно поэкспериментировать с размером шума и другими параметрами, чтобы получить нужные изображения. Готовую модель необходимо сохранить и разработать небольшой GUI (любая готовая библиотека, можно gradio).

5. **Разработать контейнер Docker.** Разработать контейнер с готовой моделью и интерфейсом.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 16. Сегментация изображений: создайте модель, которая может разделять изображение на отдельные сегменты – для обнаружения животных на камерах слежения.

Цель: обучение нейронной сети для распознавания лиц в организации.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1 **Хранение данных.** Разработать архитектуру базы данных для хранения информации о животных.

2 **Обработка данных.** Обработать данные для задачи сегментации.

3 Разработка модели. Разработать нейронную сеть (OpenCV или любую другую модель).

4 Развитие и тестирование модели. Необходимо провести тестирование модели на новых данных, оптимизировать скорость работы модели для работы с видео.

5 Разработка API. Разработать интерфейс для быстрого доступа к модели.

6 Разработать контейнер Docker. Разработать контейнер с готовой моделью.

Результат работы программного продукта

1 Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2 Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 17. Анализ временных рядов: создайте модель, которая может анализировать временные ряды данных, для прогнозирования погоды.

Цель: разработать рекомендательной модели для прогнозирования погоды.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. Сбор данных. Данные о погоде можно найти в различных источниках.

2. Очистка и хранение данных. Очистить данные от ошибок и разработать архитектуру базы данных для хранения информации о погоде.

3. Разработка модели. Разработать модель прогнозирования временного ряда.

4. Оценка и тестирование. Протестировать работу модели.

5. Интеграция модели в приложение. Создать простой веб-интерфейс, чтобы пользователи могли вводить данные и получать прогнозы.

Результат работы программного продукта.

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 18. Распознавание речи: создайте модель распознавания речи, чтобы преобразовать речь в текст, которое понимает машина.

Цель: обучение нейронной сети для распознавания речи для голосового помощника.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. Хранение данных. Разработать архитектуру базы данных для хранения информации о людях и их речи.
2. Обработка данных. Обработать данные для задачи биометрии – векторизовать звук в матрицы.
3. Разработка модели. Разработать нейронную сеть и провалидировать модель.
4. Развитие и тестирование модели. Необходимо провести тестирование модели на новых данных, оптимизировать скорость работы модели для работы с видео.
5. Разработать контейнер Docker. Разработать контейнер с готовой моделью.

Результат работы программного продукта.

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.
2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 19. Планирование маршрутов: создайте модель, которая может планировать оптимальные маршруты для транспорта или других проектов.

Цель: обучение модели для поиска оптимального маршрута.

Этапы задачи:

1. Разработать скрипт для парсинга расстояний и дорог.
2. Предобработать данные – очистить от дубликатов/пропусков или заполнить пропуски.
3. Разработать архитектуру и модель БД.
4. Обучить модель машинного обучения на основе данных – для рекомендации минимального расстояния и выбора оптимального маршрута.
5. Насытить данные для лучшей работы модели.
6. Сохранит модель для будущей работы.
7. Разработать контейнер с автоматической обработкой данных и обучением модели.
8. Разработать простой веб-интерфейс для быстрого доступа к модели.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Практико-ориентированная задача (кейс) 20. Автоматизация процессов: создайте модель, которая может автоматизировать повторяющиеся бизнес-процессы, например, обработку заказов или сбор данных.

Цель: обучение модели для оптимизации процессов бизнеса.

Этапы задачи:

1. Провести EDA – данные по работе каждого отдела и работы, KPI сотрудников.
2. Собрать полный отчет в Jupyter notebook – основные закономерности в данных и статистический анализ. Сформировать витрину данных.
3. Рассмотреть возможные способы автоматизации рутинных задач отделов.
4. Подобрать модели машинного обучения и обучить на тех задачах, которые хотим упростить.
5. Сохранит модель для будущей работы.
6. Создать презентацию с возможным решением проблем и реализации решений.
7. Запустить модель и провести A/B-тестирование.

В качестве результата работы программного продукта выложить результат на github и предоставить инструкцию.

Практико-ориентированная задача (кейс) 21. Распознавание объектов: создайте модель, которая может распознавать объекты на изображении или в реальном времени - автоматическое распознавание номеров автомобилей.

Цель: обучение нейронной сети для распознавания номеров машин.

Для разработки модели, необходимо выполнить следующие задачи.

Этапы задачи:

1. Обработка данных. Обработать данные для задачи сегментации и определения значений на номере.
2. Разработка модели. Разработать нейронную сеть (OpenCV или любую другую модель).
3. Разработка БД. Разработать архитектуру БД для хранения номеров и машин, а также изображений.

4. Развитие и тестирование модели. Необходимо провести тестирование модели на новых данных, оптимизировать скорость работы модели для работы с видео.

5. Разработка API. Разработать интерфейс для быстрого доступа к модели.

6. Разработать контейнер Docker. Разработать контейнер с готовой моделью.

Результат работы программного продукта

1. Документация, которая описывает каждый шаг работы с данными и моделью, а также их использование в реальных условиях.

2. Инструкции по установке и настройке окружения для работы с разработанным кодом и моделью в реальных условиях. Это может включать в себя создание Docker-контейнеров, установку зависимостей и прочее.

Критерии оценивания, шкала оценивания:

0–4 балла: имеются содержательные и логические ошибки, решение кейса не найдено;

5–6 баллов: решение кейса в целом найдено, но оно неоптимально и/или имеются логические ошибки;

7–8 баллов: решение кейса найдено, но имеются неточности в решении;

9–10 баллов: решение кейса найдено, ошибки отсутствуют.

Максимально возможное число баллов – 10:

не менее 9 баллов – «отлично»;

7–8 баллов – «хорошо»;

5–6 баллов – «удовлетворительно»;

0–4 балла – «неудовлетворительно».

В ходе итоговой аттестации обучающиеся должны продемонстрировать следующие знания, умения и навыки:

Знать:

– функции, задачи, навыки, содержание работы архитектора в области искусственного интеллекта;

– определения, историю развития и главные тренды искусственного интеллекта;

– основные типы, источники больших данных, проблемы безопасности;

– методы и программный инструментарий технологий больших данных;

– синтаксис и семантику основных языков искусственного интеллекта (Python, JAVA/C#/C++/Scala) и основные приемы программирования на них;

– основные алгоритмы, используемые в языках программирования;

- базовые понятия и возможности теории игр для их применения в области машинного обучения и искусственного интеллекта;
- методы, инструментальные средства и стандарты статистического и описательного анализа данных;
- системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, SPARK и др.): принципы работы, функционал, язык запросов, преимущества и недостатки;
- основные платформы данных (облачные и внутрикорпоративные), применяемые при разработке решений на основе ИИ;
- архитектуру, принципы, процесс, стадии и методологии разработки решений на основе ИИ;
- основные виды представления данных: табличные, графовые, временные ряды;
- принципы, содержание и области применения массово-параллельной обработки и анализа данных для ускорения машинного обучения;
- основы методов оптимизации;
- технологии, методы и инструменты машинного обучения для решения профессиональных задач;
- технологии распределенной обработки данных;
- принципы и содержание работы с распределенными кластерными системами;
- основные уровни представления данных;
- технологии и методы потоковой обработки данных;
- архитектуры современных нейронных сетей и их использование для решения профессиональных задач.

Уметь:

- применять системы обработки и анализа больших массивов данных;
- применять системы контроля версий Git для разработки решений на основе языков программирования;
- применять язык SQL для прикладных решений;
- применять язык программирования Python и библиотеки при разработке решений на основе ИИ;
- использовать основные статистические методы анализа данных;
- визуализировать анализируемые данные;
- проводить EDA — разведочный анализ данных;
- моделировать данные в хранилищах (DWH, Data Lake, NoSQL хранилища данных);

- применять методы машинного обучения для работы с большими данными;
- использовать шины данных (kafka) при обработке больших данных;
- строить модель для прогнозирования временного ряда;
- строить модель для рекомендательных систем;
- использовать контейнеризацию и развертывать модель в производство;
- применять технологии компьютерного зрения для обработки изображений и видео, задач биометрии;
- применять технологии обработки естественного языка для анализа и генерации текстов;
- применять технологии обучения с подкреплением для задач автоматизации производства и транспорта;
- использовать цифровые платформы анализа данных для решения профессиональных задач.

Владеть:

- языками программирования (Python, JAVA/C#/C++/Scala) для реализации методов анализа и структурирования данных;
- основными инструментами для работы с данными (Git, Docker, CI/CD, RestAPI, Jupyter notebook, kaggle, Nvidia Cuda, VS, IntelliJIdea);
- навыками работы с основными библиотеками Python для визуализации данных (Matplotlib, Seaborn, Plotly);
- навыком создания логической и физической базы данных;
- навыком установки Hadoop на виртуальную машину;
- навыком использования Docker;
- навыком работы с MapReduce, PySpark;
- навыком BI-аналитики;
- навыком работы с Numpy;
- способами улучшить качество модели с помощью методов Feature engineering;
- навыком использования библиотеки PyTorch, применения GPU в PyTorch;
- навыком построения полносвязной нейронной сети для задачи классификации;
- навыком работы с генеративными моделями GAN для задач генерации;
- навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий;
- навыком работы с технологией RL — обучения с подкреплением;
- языками запросов для обращения к СУБД;

- основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных;
- методами машинного обучения и нейронных сетей;
- математическими методами анализа данных;
- культурой постановки и планирования последовательности решения задач анализа данных и классификации;
- навыками анализа реальных задач из различных предметных областей на уровне отдельных подходов и коллективами алгоритмов;
- навыками работы с библиотеками, программными платформами (фреймворками) и программными комплексами машинного обучения.

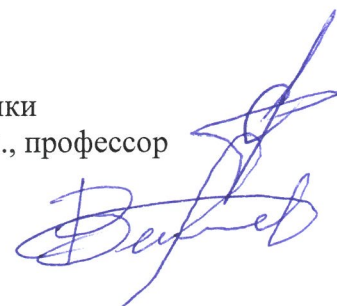
Тугой И.А. Академический директор ООО «1Т»



Разработчики программы:

Борисов В. В. Профессор кафедры вычислительной техники филиала НИУ «МЭИ» в г. Смоленске, д.т.н., профессор

Бен Режеб Тауфик Декан факультета «Радио и телевидение»
Бен Камель МТУСИ



Дроздов С.Н. Доцент Института компьютерных технологий и информационной безопасности ФГАОУ ВО «Южный федеральный университет», к.т.н.



Захарова А. А. Главный научный сотрудник ИПУ РАН, д.т.н., доцент

Ерохин К.С. Руководитель отдела Data Science ООО «1Т»

Ерохин В.А. Ведущий специалист по ИИ, ООО «1Т»

Кузин А.А. Старший преподаватель, Департамент математического и компьютерного моделирования, Институт математики и компьютерных технологий МТУСИ

Бондаренко Е.В. Ведущий специалист по ИИ, ООО «1Т»

Козлова А.С. Специалист по ИИ, ООО «1Т»

Юров Ю.А. Специалист по ИИ, ООО «1Т»

Дон Ю.Н. Разработчик машинного обучения, ООО «Рубиус тех»

Софьина К.Г. Специалист по ИИ, ООО «1Т»

