

Автономная некоммерческая организация высшего образования
«Российский новый университет» (РосНОУ)
Институт дополнительного образования



УТВЕРЖДАЮ
Проректор по качеству
образования и аккредитации
И.В. Дарда
30 марта 2023 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ
«Технический аналитик»

Срок обучения: 250 академических часов

Москва – 2023

Оглавление

1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ	4
1.1 Цель реализации программы	4
1.2 Актуальность.....	4
1.3 Нормативно-правовая основа подготовки программы	6
1.4 Требования к уровню подготовки слушателя.....	7
1.5 Область профессиональной деятельности	8
1.6. Планируемые результаты обучения.....	8
2. УЧЕБНЫЙ (ТЕМАТИЧЕСКИЙ) ПЛАН	12
3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК	15
4. УЧЕБНАЯ (РАБОЧАЯ) ПРОГРАММА МОДУЛЯ 1.....	166
5. УЧЕБНАЯ (РАБОЧАЯ) ПРОГРАММА МОДУЛЯ 2.....	25
6. ФОРМЫ АТТЕСТАЦИИ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ	37
6.1 Диагностика базовых знаний (вступительное испытание).....	37
6.2 Текущий контроль знаний	39
6.3 Промежуточная аттестация	41
6.4 Итоговая аттестация	42
7. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ	56
7.1 Специфика организационных действий и педагогических условий	56
7.2 Кадровое обеспечение программы (преподавательский состав)	56
7.3 Учебно-методическое обеспечение и информационное сопровождение	57
7.4 Материально-технические условия реализации программы.....	622

ОБЩИЕ ДАННЫЕ
О ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПРОГРАММЕ ПОВЫШЕНИЯ
КВАЛИФИКАЦИИ «ТЕХНИЧЕСКИЙ АНАЛИТИК»

№	Название	Описание
1	Название программы	Технический аналитик
2	Специальность	Технический аналитик (AI/TADS Analytic)
3	Целевое назначение	Разработка и использование технологий искусственного интеллекта
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
5	Количество академических часов	250
6	Количество слушателей, которое может обеспечить обучением Провайдер по Образовательной программе по одному потоку в срок до 5 декабря	210
7	Стоимость обучения	85 000

Разработчики программы:

1. Балькина Анна Михайловна, доцент общеуниверситетской кафедры основ математики и информатики, канд. психол. наук АНО ВО «Российский новый университет»,
2. Золотарев Олег Васильевич, заведующий кафедрой «Информационные системы в экономике и управлении» Института информационных систем и инженерно-компьютерных технологий доцент, канд. техн. наук, АНО ВО «Российский новый университет»,
3. Лабунец Леонид Витальевич, профессор кафедры «Информационные системы в экономике и управлении» Института информационных систем и инженерно-компьютерных технологий, докт. техн. наук, АНО ВО «Российский новый университет»,
4. Михалёва Елизавета Вячеславовна, заместитель исполнительного директора Института информационных систем и инженерно-компьютерных технологий, АНО ВО «Российский новый университет»,
5. Палкин Евгений Алексеевич, проректор по научной работе и заместитель председателя Учёного совета, канд. физ.-мат. наук, профессор, почетный работник высшего профессионального образования Российской Федерации, лауреат Государственной премии СССР, АНО ВО «Российский новый университет».

Дополнительная профессиональная программа повышения квалификации «Технический аналитик»

1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1 Цель реализации программы

Цель реализации программы: приобретение слушателями профессиональных компетенций в области информационных технологий в части машинного обучения и анализа больших данных по направлению подготовки «Технический аналитик».

1.2 Актуальность

Формируемый в мире и в России новый технологический уклад базируется на использовании информационно-телекоммуникационных технологий, создании и развитии информационного общества. Проникновение цифровых технологий во все сферы деятельности Российской Федерации служит причиной существенных изменений привычных моделей экономического и социального развития государства и условием его инклюзивного устойчивого развития, что и определило приоритетность тенденции цифровизации. В частности, цифровизация экономики Российской Федерации на сегодняшний день в рамках национального проекта «Национальная программа “Цифровая экономика Российской Федерации”» (утв. президиумом Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам, протокол от 04.06.2019 № 7) определена одним из приоритетных направлений государственной политики и рассматривается в качестве инновационной составляющей конкурентного поступательного движения и крупномасштабных качественных преобразований, центрального звена стратегического управления и фундаментального компонента экономического суверенитета государства.

Промышленность Российской Федерации может стать драйвером цифровой экономики и информационного общества в стране, сформировав необходимый базис для обеспечения технологического суверенитета, что в свою очередь требует следования концепции устойчивого развития, промышленной революции, циркулярной и цифровой экономики.

Интеграция физических и цифровых ресурсов в контексте цифровизации выступает основным средством трансформации системы образования, которая посредством использования инновационных цифровых методов, технологий и инструментов направлена на актуализацию образовательного процесса в аспекте соответствия мировым тенденциям информатизации общества для повышения качественного уровня профессиональной подготовки специалистов в стремительно меняющейся и продолжающей свое формирование цифровой экономике. Этому процессу в образовательной среде присущи форсированная автоматизация и роботизация производства и управления, коренные изменения социальных структур – расширение сферы информационной деятельности и услуг, что в совокупности, согласно Указу Президента Российской Федерации от 7 мая 2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года», является одной из национальных целей развития Российской Федерации на период до 2024 года.

Информационное общество определено как общество, в котором информация и уровень ее применения и доступности кардинальным образом влияют на экономические и социокультурные условия жизни граждан. Существенным этапом на пути развития информационного общества в Российской Федерации стало принятие ряда значимых документов, в частности, Указа Президента Российской Федерации «О стратегии развития информационного общества в РФ на 2017-2030 годы», создание Автономной некоммерческой организации «Цифровая экономика», миссией которой является создание экосистемно-центричной экономики для удобства жизни граждан и ведения бизнеса в России, введение в действие в 2019 году плана информатизации Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации. В дальнейшем, в 2020 году, цифровизация экономики была простимулирована пандемией коронавируса, которая сопровождалась вынужденным режимом самоизоляции граждан, что повлекло массовый переход на дистанционные формы взаимодействия, требующие активного применения цифровых технологий и соответствующей инфраструктуры. В этом же году цифровизация была объявлена национальной целью развития России.

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.

Под искусственным интеллектом в Программе развития в соответствии с «Национальной стратегией развития искусственного интеллекта на период до 2030 года», утверждённой 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.

Федеральный проект «Искусственный интеллект» предусматривает повышение уровня кадрового обеспечения российского рынка технологий искусственного интеллекта, внедрение инновационных решений в разные сферы хозяйствования, повышение доступности и качества данных, необходимых для развития технологий искусственного интеллекта.

Следовательно, первоочередным на нынешнем этапе является формирование принципиально новых направлений деятельности во всех отраслях экономики на основе технологий искусственного интеллекта, что предполагает одновременно дальнейшее развитие целостной системы устойчивого воспроизводства и привлечения кадров при тесном сотрудничестве и взаимодействии научных и образовательных организаций, предпринимательского сообщества, государственных корпораций и компаний, создание новых высокопроизводительных рабочих мест и повышение уровня занятости населения, оптимизация процессов подбора и обучения кадров. В контексте повышения занятости населения и развития кадрового потенциала в 2022 году осуществлен запуск масштабного проекта для обеспечения получения гражданами дополнительного профессионального образования в области искусственного интеллекта и в смежных областях с использованием механизма персональных цифровых сертификатов.

В современных реалиях информационного общества данные – это, прежде всего, «кровь» предприятия, то, без чего невозможно жизнеобеспечение и существование любой компании. Поэтому в любой сфере национальной экономики функционирование организации должно быть реализовано на основе гибкого управления данными, в том числе большими (Big Data), и оно станет успешным, если все её бизнес-процессы и организационная структура построены на основе сквозной интеграции информационных потоков, их непрерывной и прогнозной предиктивной аналитики.

Успешную деятельность любой организации в условиях растущих объемов информации сегодня невозможно представить без такого профессионала, как технический аналитик.

Технический аналитик (бизнес аналитик) – это специалист, который исследует финансовый рынок с помощью инструментов технического анализа и строит прогноз на основе его результатов. Одной из ключевых задач бизнес-аналитика является выявление проблем бизнеса и поиск максимально эффективного решения в каждой конкретной ситуации. Для этого он должен обладать соответствующими знаниями в предметной области.

Актуальность программы обусловлена необходимостью подготовки квалифицированных кадров в условиях нарастающих требований цифровизации к постоянному совершенствованию их профессиональных, цифровых и социальных навыков.

1.3 Нормативно-правовая основа подготовки программы

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

2. Приказ Министерства образования и науки Российской Федерации от 01 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам».

3. Приказ Министерства образования и науки Российской Федерации от 23 августа 2017 г. № 816 «Об утверждении порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ».

Дополнительная профессиональная программа (программа повышения квалификации) «Технический аналитик» (далее – Программа) разработана:

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 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» (далее – приказ Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации № 143);

с учетом федеральных государственных образовательных стандартов (далее вместе – ФГОС ВО) по направлению подготовки 01.03.02 Прикладная математика и информатика (Приказ Минобрнауки России от 10.01.2018 N 9 (ред. от 08.02.2021), 09.03.02 Информационные системы и технологии (Приказ Минобрнауки России от 19.09.2017 N 926 (ред. От 08.02.2021));

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

1.4 Требования к уровню подготовки слушателя

Программа рассчитана на специалистов IT-организаций с высшим или средним профессиональным образованием, а также работников из других профессиональных областей, желающих получить новые компетенции по специальности «Технический аналитик».

Также к обучению допускаются студенты высших и средних профессиональных образовательных организаций.

Требуются углубленные знания математики, алгоритмов и структур данных, продвинутые навыки программирования на любом языке.

Наличие опыта профессиональной деятельности: без опыта.

Требования к уровню подготовленности потенциальных слушателей.

Слушатели должны обладать следующими знаниями, умениями и владеть навыками:

ЯЗЫКИ ПРОГРАММИРОВАНИЯ

- Умение писать программы на каком-либо императивном языке программирования (Java, Python, JavaScript, C++ и т.п.)
- Понимание базовых структур данных
- Владение основами ООП (класс, объект)
- Базовые знания алгоритмов, структур и баз данных

АНАЛИТИКА

- Понимание базовых основ математической статистики
- Понимание базовых основ теории вероятностей
- Базовые знания в выборе необходимого метода визуализации данных

Требования к профессиональным компетенциям 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.

1.5 Область профессиональной деятельности

06 Связь и информационно-коммуникационные технологии (предпринимательская деятельности в области информационных технологий, проектно-исследовательская деятельность в области информационных технологий, автоматизация информационно-аналитической деятельности, создание и применение технологии больших данных, разработка компьютерного программного обеспечения; деятельность по обработке данных, предоставление услуг по размещению информации и связанная с этим деятельность).

1.6. Планируемые результаты обучения

Программа повышения квалификации разработана с учетом профессиональных стандартов 06.042 Специалист по большим данным, 06.022 Системный аналитик, 08.037 Бизнес-аналитик.

Программа повышения квалификации разработана с учетом: ФГОС 38.03.05 Бизнес-информатика, ФГОС 09.04.04. Программная инженерия.

Программа направлена на совершенствование следующих профессиональных компетенций (далее – ПК) из универсальной модели компетенций в сфере искусственного интеллекта, разработанной РЭУ им. Г.В. Плеханова в 2021 году в рамках результата Федерального проекта «Искусственный интеллект»:

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

ПК-4.р. – способность разрабатывать и применять методы машинного обучения для решения задач;

ПК-5.р. – способность использовать инструментальные средства для решения задач машинного обучения;

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

ПК-7.п – способность выполнять анализ больших данных;

ПК-7.р – способность осуществлять сбор и подготовку данных для систем искусственного интеллекта;

ПК-8.и. – способность использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта;

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

Компетенция	Критерии оценки уровня сформированности компетенций
ПК-1.р. 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. Собирает исходную информацию и формирует требования к решению задач с использованием методов искусственного интеллекта.

ПК-4.р. Способен разрабатывать и применять методы машинного обучения для решения задач	Проводит анализ требований и определяет необходимые классы задач машинного обучения. Определяет метрики оценки результатов моделирования и критерии качества построенных моделей. Принимает участие в оценке и выборе используемых методов машинного обучения.
ПК-5.р. Способен использовать инструментальные средства для решения задач машинного обучения	Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи. Разрабатывает модели машинного обучения для решения задач.
ПК-6.р. 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. Разрабатывает системы искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств.
ПК-7.п. Способен выполнять анализ больших данных	Использует знания о вариантах использования больших данных, определениях, словарях и эталонной архитектуре больших данных для эффективного извлечения, хранения, подготовки больших данных.
ПК-7.р. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения.
ПК-8.и. Способен использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	Решает прикладные задачи и участвует в реализации проектов в области сквозной цифровой субтехнологии «Обработка естественного языка».
ПК-9.р. Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	Участвует в реализации проектов в области сквозной цифровой субтехнологии «Обработка естественного языка».

В результате освоения данной программы слушатель программы «Технический аналитик» должен:

Знать:

- базовые понятия искусственного интеллекта и BigData;
- отличительные особенности машинного обучения, нейронных сетей;
- основы программирования на языке Python;
- ключевые компоненты и парадигмы Hadoop;
- основы теории алгоритмов и структур данных;
- подходы и принципы работы с базами данных NoSQL;

- приемы и принципы управления СУБД на основе языка запросов SQL;
- базовые модели и уровни представления данных;
- основные функции и инструменты процесса ETL;
- преимущества организации MPP- СУБД для хранения и аналитической обработки больших данных;
- особенности и архитектура СУБД Postgres;
- принципы действия, преимущества и область применения колоночного формата хранения данных;
- правила и принципы создания отчетов в контексте визуализации данных;
- фундаментальные основы теории вероятности;
- основные положения комбинаторики;
- суть метода и базовые инструменты A/B-тестирования;
- подходы, принципы и инструменты продуктовой аналитики;
- аналитический инструментарий и перспективные методики технологий больших данных.
- классические модели машинного обучения и популярные архитектуры нейронных сетей.

Уметь:

- анализировать широкий спектр информации и данных с целью повышения эффективности конкретных ресурсов/процессов или внедрения решений по автоматизации;
- осуществлять решение прикладных задач с использованием SQL базы данных;
- разрабатывать алгоритмы прикладных решений на основе ИИ с использованием языка программирования и библиотек Python;
- применять современные методы и инструменты поиска и структурирования данных;
- применять технику визуализации для интеллектуального анализа данных;
- применять принципы интеллектуального анализа данных;
- применять графовые модели анализа данных;
- выполнять исследования и расчеты показателей проекта на основе методов имитационного моделирования вероятностных распределений с использованием языка программирования Python;
- проводить исследования поискового и аналитического характера с помощью инструментов технического анализа на основе методов и техник анализа больших данных;
- создавать и внедрять математические и информационные модели в зависимости от целей аналитических исследований;
- проектировать технические решения проектов в сфере информационных технологий и обеспечивать их поддержку на протяжении всего жизненного цикла.

Владеть:

- навыками решения ситуационных аналитических задач с использованием алгоритмов искусственного интеллекта и инструментов визуализации для создания интерактивных дашбордов;
- статистическими методами анализа данных;
- навыками разведочного, описательного и исследовательского анализа данных с использованием языка Python и его инструментальных средств для расчета показателей;
- методами построения моделей машинного обучения и нейронных сетей;
- математическими методами анализа и обработки технологической информации;
- навыками разработки таблиц в СУБД Postgres и управления ими с помощью Dbeaver;

- навыками создания полносвязной нейронной сети для решения задач классификации на основе любых наборов данных;
- навыками использования фреймворков Scikit-learn, Scipy для обучения моделей машинного обучения;
- навыками использования фреймворков PyTorch, TensorFlow и Keras для обучения нейронных сетей;
- навыками анализа финансовых и продуктовых метрик и поиска «слабых» мест с помощью Python.

Выдаваемый документ при успешном освоении программы: удостоверение о повышении квалификации АНО ВО «Российский новый университет».

2. УЧЕБНЫЙ (ТЕМАТИЧЕСКИЙ) ПЛАН

№ п/п	Наименование модулей/тем программы	Всего, час	Виды учебных занятий			Формы контроля
			Лекции	Практические занятия	Самостоятельная работа	
1	Модуль 1. Основной	110	28	54	24	
2	Раздел 1. Введение в аналитику данных	20	6	8	6	
3	Тема 1.1 Введение в аналитику данных. Наука о данных, анализ данных и машинное обучение	8	2	4	2	Тест + Практическая работа
4	Тема 1.2 Анализ данных в разных секторах. Современные тенденции в анализе данных	6	2	2	2	Тест + Практическая работа
5	Тема 1.3 Введение в большие данные и искусственный интеллект	6	2	2	2	Тест + Практическая работа
6	Раздел 2. Основы статистики	36	10	18	8	
7	Тема 2.1 Введение в теорию вероятностей. Базовые понятия	8	2	4	2	Тест + практическая работа
8	Тема 2.2 Введение в математическую статистику. Статистические методы анализа данных	10	2	6	2	Тест + практическая работа
9	Тема 2.3 Проверка статистических гипотез	8	2	4	2	Тест + практическая работа
10	Тема 2.4 A/B-тестирование	10	4	4	2	Тест + практическая работа
11	Раздел 3. Инструменты и технологии анализа данных	50	12	28	10	
12	Тема 3.1 Основы программирования на Python. Знакомство с средой разработки Jupyter Notebook. Введение в язык R	10	2	6	2	Тест + Практическая работа
13	Тема 3.2 Введение в основные	10	2	6	2	Практическая работа

	библиотеки Python для анализа					
14	Тема 3.3 Введение в визуализацию данных. Python-библиотеки Matplotlib, Seaborn, Plotly	10	2	6	2	
15	Тема 3.4 Парсинг интернет-данных. Работа с API и http	8	2	4	2	Тест + Практическая работа
16	Тема 3.5. Обзор систем обработки и анализа больших данных Hadoop, ETL, Spark	12	4	6	2	Тест + Практическая работа
17	<i>Промежуточная аттестация</i>	4				
18	Модуль 2. Профилирующий	132	38	56	34	
19	Раздел 4. Технологии хранения и обработки больших данных	54	18	22	14	
20	Тема 4.1 Введение в теорию БД. Моделирование данных. Основы языка SQL	10	4	4	2	Тест + практическая работа
21	Тема 4.2 Реляционные базы данных. Языки запросов реляционных СУБД. Декларативный язык SQL	8	2	4	2	Тест + Практическая работа
22	Тема: 4.3 NoSQL хранилища данных	6	2	2	2	Тест + Практическая работа
23	Тема 4.4 Анализ данных при массово параллельной обработке. Обзор GreenPlum	4	2	0	2	Тест + Практическая работа
24	Тема 4.5 Знакомство с СУБД Postgres. Расширения SQL в PostgreSQL	8	2	4	2	
25	Тема 4.6 Введение в СУБД Clickhouse	6	2	2	2	Тест + Практическая работа
26	Тема 4.7 BI инструменты. Построение дашбордов с Apache Superset	12	4	6	2	Тест + Практическая работа
27	Раздел 5. Машинное обучение и оптимизация	30	8	14	8	
28	Тема 5.1 Математические основы машинного обучения	8	2	4	2	Тест + Практическая работа

29	Тема 5.2 Обзор методов анализа больших данных. Выбор и обработка данных для машинного обучения	8	2	4	2	Тест + Практическая работа
30	Тема 5.3 Построение моделей машинного обучения	8	2	4	2	Тест + Практическая работа
31	Тема 5.4 Инструменты анализа данных и машинного обучения (Rapid Miner)	6	2	2	2	Тест + Практическая работа
32	Раздел 6. Нейронные сети	32	8	16	8	
33	Тема 6.1 Введение в нейронные сети. Глубокое обучение	8	2	4	2	Тест + Практическая работа
34	Тема 6.2 Архитектура нейронных сетей. Сверточные и рекуррентные сети	8	2	4	2	Тест + Практическая работа
35	Тема 6.3 Обучение нейронных сетей с помощью TensorFlow и Keras	8	2	4	2	Тест + Практическая работа
36	Тема 6.4 Анализ естественного языка с помощью методов искусственного интеллекта	8	2	4	2	Тест + Практическая работа
37	Раздел 7. Продвижение продукта. Бизнес-метрики	12	4	4	4	
38	Тема 7.1. Введение в продуктовую аналитику	6	2	2	2	Тест + практическая работа
39	Тема 7.2. Ключевые метрики роста продукта	6	2	2	2	Тест + практическая работа
40	<i>Промежуточная аттестация</i>	4				
41	Итоговая аттестация	8				
	Всего часов	250	66	110	58	

3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

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

Продолжительность обучения – 3 месяца

Режим занятий: 3-4 часа в день.

Завершение обучения: не позднее 05.12.2023

№№	Учебные недели	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Модуль 1. Основной Раздел 1. Введение в аналитику данных	Л П Х													
2.	Модуль 1. Раздел 2. Основы статистики		Л П Х	Л П Х											
3.	Модуль 1. Раздел 3. Инструменты и технологии анализа данных				Л П Х	Л П Х									
4.	Промежуточная аттестация					Л П Х									
5.	Модуль 2. Профилирующий. Раздел 4. Технологии хранения и обработки больших данных						Л П Х	Л П Х	Л П Х						
6.	Модуль 2. Раздел 5. Машинное обучение и оптимизация									Л П Х	Л П Х				
7.	Модуль 2. Раздел 6. Нейронные сети											Л П Х	Л П Х		
8.	Модуль 2. Раздел 7. Продвижение продукта. Бизнес-метрики														Л П Х
9.	Промежуточная аттестация														
10.	Итоговая аттестация														Л П Х ИА

Условные обозначения: Л-лекции, П – практические занятия, Х – самостоятельная работа, ИА – итоговая аттестация.

4. УЧЕБНАЯ (РАБОЧАЯ) ПРОГРАММА МОДУЛЯ 1

Цель освоения модуля: приобретение слушателями профессиональных компетенций в области анализа данных и статистики, инструментов для анализа больших данных.

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

ПК-7.п. — способность выполнять анализ больших данных;

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

Планируемые результаты обучения по модулю 1:

Знать:

- основы статистического анализа;
- статистические методы;
- типы больших данных;
- источники данных;
- методы проведения анализа данных;
- суть метода и базовые инструменты А/В-тестирования.

Уметь:

- применять современные методы и инструменты поиска и структурирования данных;
- применять технику визуализации для интеллектуального анализа данных;
- применять принципы интеллектуального анализа данных;
- применять графовые модели анализа данных;

Владеть:

- статистическими методами анализа данных;
- навыками использования технологии хранения и обработки больших данных в организации;
- технологиями анализа данных.

Учебно-тематический план модуля 1

п/п	Наименование модулей/ тем программы	Всего, час	Виды учебных занятий			Формы контроля
			лекции	практические занятия	самостоятельная работа	
1	Модуль 1. Основной	110	28	54	24	
2	Раздел 1. Введение в аналитику данных	20	6	8	6	
3	Тема 1.1 Введение в аналитику данных. Наука о данных, анализ данных и машинное обучение	8	2	4	2	Тест + Практическая работа

4	Тема 1.2 Анализ данных в разных секторах. Современные тенденции в анализе данных	6	2	2	2	Тест + Практическая работа
5	Тема 1.3 Введение в большие данные и искусственный интеллект	6	2	2	2	Тест + Практическая работа
6	Раздел 2. Основы статистики	36	10	18	8	
7	Тема 2.1 Введение в теорию вероятностей. Базовые понятия	8	2	4	2	Тест + практическая работа
8	Тема 2.2 Введение в математическую статистику. Статистические методы анализа данных	10	2	6	2	Тест + практическая работа
9	Тема 2.3 Проверка статистических гипотез	8	2	4	2	Тест + практическая работа
10	Тема 2.4 А/В-тестирование	10	4	4	2	Тест + практическая работа
11	Раздел 3. Инструменты и технологии анализа данных	50	12	28	10	
12	Тема 3.1 Основы программирования на Python. Знакомство с средой разработки Jupyter Notebook. Введение в язык R	10	2	6	2	Тест + Практическая работа
13	Тема 3.2 Введение в основные библиотеки Python для анализа	10	2	6	2	Практическая работа
14	Тема 3.3 Введение в визуализацию данных. Python-библиотеки	10	2	6	2	

	Matplotlib, Seaborn, Plotly.					
15	Тема 3.4 Парсинг интернет-данных. Работа с API и http.	8	2	4	2	Тест + Практическая работа
16	Тема 4. Обзор систем обработки и анализа больших данных Hadoop, ETL, Spark.	12	4	6	2	Тест + Практическая работа
17	Промежуточная аттестация	4		4		Решение кейса

Содержание модуля 1

Наименование учебных модулей и тем	Виды учебных занятий	Содержание учебных занятий
Модуль 1. Основной		
Раздел 1. Введение в аналитику данных		
Тема 1.1 Введение в аналитику данных. Наука о данных, анализ данных и машинное обучение	Лекция (2 ч)	Введение в аналитику данных. Наука о данных, анализ данных и машинное обучение. Работа с разными типами данных. Методология обработки данных Технический аналитик: потребность и ценность.
	Практическое занятие (4 ч)	1. Решить кейс по анализу условного продукта на базе имеющихся знаний. Предложить список необходимых данных для анализа и продумать ключевые метрики. 2. Провести анализ рынка для «Технического аналитика» в РФ.
	Самостоятельная работа (2 ч)	Самостоятельно разобрать типовые кейсы, которые решают технические аналитики данных в реальной практике. Тест 1.1: Введение в аналитику данных.
Тема 1.2 Анализ данных в разных секторах. Современные тенденции в анализе данных	Лекция (2 ч)	Анализ больших данных: будущее за Big Data. Анализ данных и научный подход: как новые технологии меняют промышленный сектор. Современные тенденции в анализе данных. Рост AI Ops. Демократизация данных. Интеграция данных. Малые и широкие данные.
	Практическое занятие (2 ч)	Предложить интервал заработной платы на основе данных о вакансии, без использования машинного обучения. На основе подготовленной обучающей выборки из вакансий с указанными зарплатами (с верхней и нижней границей или с одной из

		них). Задача – максимально точно предположить возможную предлагаемую заработную плату, которую мог бы поставить работодатель.
	Самостоятельная работа (2 ч)	Изучить материал: Современные тенденции в мире данных и аналитики. Тест 1.2: Анализ данных в разных секторах. Современные тенденции в анализе данных
Тема 1.3 Введение в большие данные и искусственный интеллект	Лекция (2 ч)	Big Data и искусственный интеллект: технологии будущего. Основные определения ИИ и BigData
	Практическое занятие (2 ч)	Произвести аналитику для интеллектуального отслеживания ресурсов/процессов
	Самостоятельная работа (2 ч)	Изучить материал: Различия между Data Science, машинным обучением, ИИ, глубоким обучением и Data Mining Тест 1.3: Введение в большие данные и искусственный интеллект
Раздел 2. Основы статистики		
Тема 2.1 Введение в теорию вероятностей. Базовые понятия	Лекция (2 ч)	Введение в теорию вероятностей. Разбор основных понятий теории вероятностей: случайные события, алгебра событий, вероятность, условная вероятность, зависимые и независимые события, формула полной вероятности и теорема Байеса, статистическое определение вероятности
	Практическое занятие (4 ч)	Произвести расчет вероятностных показателей с использованием языка Python. Модуль random  teoria_veroyatnosti_python.ipynb
	Самостоятельная работа (2 ч)	Изучить материал: Основы комбинаторики. Изучение модуля math, itertools. Тест 2.1 Введение в теорию вероятностей
Тема 2.2 Введение в математическую статистику. Статистические методы анализа данных	Лекция (2 ч)	Введение в математическую статистику. Основные понятия: выборка, генеральная совокупность, вариационный ряд, выборочное среднее, мода, медиана, среднеквадратичное отклонение, дисперсия и т.д. Статистические методы исследования: Описательная статистика, статистические критерии Крамера-Уэлча, Вилкоксона-Манна-Уитни, хи-квадрат, Фишера и др. Корреляционный анализ, дисперсионный анализ, регрессионный анализ

	Практическое занятие (6 ч)	Практика расчета статистических показателей с использованием языка Python. Введение в модуль scipy  statistical-analysis-py thon.ipynb
	Самостоятельная работа (2 ч)	Изучить материал: Проверка статистических гипотез. Тест 2.2: Статистические методы анализа данных
Тема 2.3 Проверка статистических гипотез	Лекция (2 ч)	Общий обзор. Определение нулевой и альтернативной гипотезы, уровня статистической значимости. Получение статистики критерия, определение критической области. Получение значения p (достигнутого уровня значимости). Применение значения p . Проверка гипотез против доверительных интервалов
	Практическое занятие (4 ч)	Решить задачу: Проверка гипотезы о среднем значении веса чая. Продавец утверждает, что средний вес пачки чая составляет 100 г. Из партии извлечена выборка и взвешена. Вес каждой пачки – см. таблицу вариантов. Не противоречит ли это утверждению продавца? Доверительный интервал 99%. Вес пачек чая распределен нормально. Выборка: 98, 104, 97, 97, 101, 100, 99, 101, 99, 98
	Самостоятельная работа (2 ч)	Изучить материал: Альтернативная гипотеза, альтернатива. Тест 2.3: Проверка статистических гипотез
Тема 2.4 А/В-тестирование	Лекция (4 ч)	Понятие А/В-тестирования. Где и зачем применяется. Критерии оценки качества проведенного А/В-тестирования. Математические основы А/В-тестирования
	Практическое занятие (4 ч)	А/В-тестирование с помощью языка Python  AB testing with Python.ipynb.txt
	Самостоятельная работа (2 ч)	Изучить материал: Альтернативные способы тестирования гипотез. Практика методов тестирования с помощью языка Python. Тест 2.4: А/В-тестирование
Раздел 3. Инструменты и технологии анализа данных		
Тема 3.1 Основы программирования на Python. Знакомство со средой разработки Jupyter	Лекция (2 ч)	Введение в программирование на языке Python. Основные конструкции языка. Переменные, объекты, типы данных. Базовые структуры данных. Основные

Notebook. Введение в язык R		операторы и управляющие конструкции – операторы условий, циклы и т. д. Знакомство со средой разработки Jupyter Notebook. Введение в язык R
	Практическое занятие (6 ч)	Установка интерпретатора Python. Запуск простейших программ на языке Python. Установка Jupyter Notebook. Решение задач на проработку базовых навыков программирования на языке Python  python_start.ipynb
	Самостоятельная работа (2 ч)	Изучить материал: синтаксис языка Python и его основных конструкций. Тест 3.1: Основы программирования на Python
Тема 3.2 Введение в основные библиотеки Python для анализа	Лекция (2 ч)	Знакомство с библиотекой для анализа данных Pandas. Работа с файлами формата xls и csv. Понятие датасета. Базовые операции работы с табличными данными
	Практическое занятие (6 ч)	Практика использования возможностей библиотеки Pandas на примере данных о поездках в такси  taxi.csv
	Самостоятельная работа (2 ч)	Изучить материал: Знакомство с библиотекой NumPy. Изучение различных форматов хранения данных: json, tsv, avro, xml. Тест 3.2: Введение в основные библиотеки Python
Тема 3.3 Введение в визуализацию данных. Python-библиотеки Matplotlib, Seaborn, Plotly.	Лекция (2 ч)	Основные формы визуализации данных: таблицы, диаграмма, гистограмма, график. Знакомство с библиотеками для визуализации данных языка Python: matplotlib и seaborn, Plotly. Решения базовых аналитических кейсов с использованием инструментов визуализации
	Практическое занятие (6 ч)	1. Решить практические задачи на Python, используя библиотеки Matplotlib, seaborn, plotly. Загрузить датасет, сделать визуализацию разных признаков, их распределения, корреляции и взаимосвязей. 2. Визуализация данных о поездках в такси для решения аналитических задач поиска аномалий, стандартного отклонения и кластеризации

		 taxi.csv
	Самостоятельная работа (2 ч)	<p>Рассмотреть примеры: Методы анализа на графах. Случайные графы, безмасштабные графы, социальные сети – сети тесного мира.</p> <p>Тест 3.3 Введение в визуализацию данных</p>
Тема 3.4 Парсинг интернет-данных. Работа с API и http.	Лекция (2 ч)	<p>Рассмотрение возможностей языка Python для получения данных из внешних API.</p> <p>Краткое введение в протоколы http и rest. Погружение в библиотеку requests. Введение в понятие парсинга и веб-скрейпинга. Разбор библиотеки lxml.</p>
	Практическое занятие (4 ч)	<p>Разработка скрипта для получения данных из открытого источника с помощью подключения к API. Разработка дополнительного скрипта для парсинга веб-страницы</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  HTMLParsing.ipynb.tx t </div> <div style="text-align: center;">  API-Scraping.ipynb.tx t </div> </div>
	Самостоятельная работа (2 ч)	<p>Изучить материал: Дополнительные возможности библиотек requests и lxml. Знакомство с другими решениями для подключения к API и парсинга веб-ресурсов – BeautifulSoup, Selenium и другими. Поиск и попытка получения данных из альтернативных открытых источников.</p> <p>Тест 3.4: Парсинг интернет-данных. Работа с API и http</p>
Тема 3.5 Обзор систем обработки и анализа больших данных Hadoop, ETL, Spark	Лекция (4 ч)	<p>Основы Hadoop. Архитектура Apache Hadoop. Apache Spark. Использование Hadoop. Создание масштабируемых решений для обработки данных с помощью Apache Hive и Apache Spark. Работа с распределенной кластерной системой. ETL процессы и инструменты</p>
	Практическое занятие (6 ч)	<p>Настройка кластера Apache Spark и Hive на Hadoop</p>
	Самостоятельная работа (2 ч)	<p>Изучить материал: Бакетирование vs партиционирование в Apache Hive и Spark.</p> <p>Тест 3.5: Обзор систем обработки и анализа больших данных Hadoop, ETL, Spark</p>
Промежуточная аттестация	4 ч	Решение кейса

Содержание самостоятельной работы слушателей:

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

Вопросы для самостоятельной работы в рамках подготовки к промежуточной аттестации:

- синтаксис языка Python и его основных конструкций;
- библиотека NumPy;
- форматы хранения данных: json, tsv, avro, xml;
- методы анализа на графах;
- случайные графы, безмасштабные графы, социальные сети – сети тесного мира;
- возможности библиотек requests и lxml;
- решения для подключения к API и парсинга веб-ресурсов – BeautifulSoup, Selenium и другие;
- поиск и попытка получения данных из альтернативных открытых источников;
- бакетирование vs партиционирование в Apache Hive и Spark.

Учебно-методическое сопровождение

Обучающие материалы представлены в виде видеолекций, текстовых и графических материалов, размещенных на платформе sprint.1T.

Список литературы:

1. Деревянко М. Э., Нилова Н. М. Обзор современных информационных систем управления бизнес-процессами. – Краснодар: ФГБОУ ВО «Кубанский государственный аграрный университет им. И. Т. Трубилина», 2021. – 497 с.
2. Джон Д. Келлехер, Брайан Мак-Нейми, Аоифе д'Арси. Основы машинного обучения для аналитического прогнозирования: алгоритмы, рабочие примеры и тематические исследования / Пер. с англ. Д. Ключин. – М.: Диалектика (Вильямс), 2019. – 656 с.
3. Дэви Силен, Арно Мейсман, Мохамед Али. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.
4. Еременко К. Работа с данными в любой сфере: как выйти на новый уровень, используя аналитику. – М.: Альпина Паблишер, 2019. – 304 с.
5. Иванова В., Перерва А. Путь аналитика. Практическое руководство IT-специалиста, 2-е изд. – СПб.: Питер, 2015. – 304 с.
6. Каменнова, М. С. Моделирование бизнес-процессов в 2 ч. Часть 2: Учебник и практикум для бакалавриата и магистратуры / Машков И. В., Крохин В. В. – М.: Юрайт, 2019. – 228 с.
7. Карпова И. П. Базы данных. Учебное пособие. – СПб.: Питер, 2021. – 240 с.
8. Ковалев С. М., Ковалев В. М. Настольная книга аналитика. Практическое руководство по проектированию бизнес-процессов и организационной структуры: Практическое руководство. – М.: 1С-Паблишинг, 2020. – 360 с.
9. Конвински Энди, Венделл Патрик, Захария Матей, Карау Холден. Изучаем Spark. Молниеносный анализ данных / Пер. с англ. М.: ДМК Пресс, 2015. – 304 с.
10. Косников С. Н. Математические методы в экономике: учебное пособие для вузов, 2-е изд., испр. и доп. – М.: Юрайт, 2019. – 172 с.
11. Кулик Б.А. Логика и математика: просто о сложных методах логического анализа. – СПб.: Политехника, 2020. – 141 с.

12. Миронов В. Профессия «бизнес-аналитик» Краткое пособие для начинающих. – М.: Альпина Паблишер, 2021. – 192 с.
13. Нархид Ния, Шапира Гвен, Палино Тодд. Apache Kafka. Поточковая обработка и анализ данных / Пер. с англ. И. Пальти. СПб.: Питер, 2021. – 320 с.
14. Поляков В. М., Агаларов З. С. Методы оптимизации: Учебное пособие. – М.: Дашков и К., 2022. – 86 с.
15. Системный анализ: учебник и практикум для вузов / В.В. Кузнецов [и др.]; под общей редакцией В.В. Кузнецова. – М.: Юрайт, 2023. – 270 с.
16. Цветков А. А. Теория и практика бизнес-анализа в ИТ. – М.: Директ-Медиа, 2020. – 150 с.

Интернет-ресурсы:

1. 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 25.03.2023)
2. Бесплатные материалы по Data Engineering от преподавателей МФТИ // Московский физико-технический институт (национальный исследовательский университет). – URL: <https://fpmi-edu.ru/free-de> (дата обращения 25.03.2023)

Описание системы оценки качества освоения модуля.

Контроль результатов освоения модуля осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации. Описание оценочных средств, форм и методов контроля представлены в разделе «Формы аттестации и оценочные материалы» настоящей программы.

5. УЧЕБНАЯ (РАБОЧАЯ) ПРОГРАММА МОДУЛЯ 2

Цель освоения модуля: изучение технологий хранения и обработки больших данных, технологий машинного обучения, оптимизации и искусственного интеллекта, основ продвижения продукта.

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

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

ПК-4.р. — способность разрабатывать и применять методы машинного обучения для решения задач;

ПК-5.р. – способность использовать инструментальные средства для решения задач машинного обучения;

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

ПК-7.р. – способность осуществлять сбор и подготовку данных для систем искусственного интеллекта;

ПК-8.и. – способность использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта;

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

Планируемые результаты обучения по модулю 2:

Знать:

- базовые понятия искусственного интеллекта;
- отличительные особенности машинного обучения, нейронных сетей, глубокого обучения и EDA;
- основы программирования на языке Python;
- ключевые компоненты и парадигмы Hadoop;
- подходы и принципы работы с базами данных NoSQL;
- приемы и принципы управления СУБД на основе языка запросов SQL;
- основные функции и инструменты процесса ETL;
- преимущества организации MPP-СУБД для хранения и аналитической обработки больших данных;
- особенности и архитектура СУБД Postgres;
- принципы действия, преимущества и область применения колоночного формата хранения данных;
- правила и принципы создания отчетов в контексте визуализации данных;
- подходы, принципы и инструменты продуктовой аналитики;
- классические модели машинного обучения и популярные архитектуры нейронных сетей

Уметь:

- осуществлять решение прикладных задач с использованием SQL-базы данных;
- разрабатывать алгоритмы прикладных решений на основе ИИ с использованием языка программирования и библиотек Python;
- применять современные методы и инструменты поиска и структурирования данных;
- применять технику визуализации для интеллектуального анализа данных;
- применять принципы интеллектуального анализа данных;
- применять графовые модели анализа данных;
- выполнять исследования и расчеты показателей проекта на основе методов имитационного моделирования вероятностных распределений с использованием языка программирования Python;

- проводить исследования поискового и аналитического характера с помощью инструментов технического анализа на основе методов и техник анализа больших данных;
- создавать и внедрять математические и информационные модели в зависимости от целей аналитических исследований;
- проектировать технические решения проектов в сфере информационных технологий и обеспечивать их поддержку на протяжении всего жизненного цикла.

Владеть:

- навыками решения ситуационных аналитических задач с использованием алгоритмов искусственного интеллекта и инструментов визуализации для создания интерактивных дашбордов;
- навыками разведочного, описательного и исследовательского анализа данных с использованием языка Python и его инструментальных средств для расчета показателей;
- методами построения моделей машинного обучения и нейронных сетей;
- математическими методами анализа и обработки технологической информации;
- навыками разработки таблиц в СУБД Postgres и управления ими с помощью Dbeaver;
- навыками создания полносвязной нейронной сети для решения задач классификации;
- навыками использования фреймворков Scikit-learn, Scipy для обучения моделей машинного обучения;
- навыками использования фреймворков PyTorch, TensorFlow и Keras для глубокого и машинного обучения нейронных сетей;
- навыками анализа финансовых и продуктовых метрик и поиска «слабых» мест с помощью Python.

Учебно-тематический план модуля 2

№ п/п	Наименование модулей/тем программы	Всего, час	Виды учебных занятий			Формы контроля
			Лекции	Практические занятия	Самостоятельная работа	
1	Модуль 2. Профилирующий	132	38	56	34	4
2	Раздел 4. Технологии хранения и обработки больших данных	54	18	22	14	0
3	Тема 4.1 Введение в теорию БД. Моделирование данных. Основы языка SQL	10	4	4	2	Тест + практическая работа
4	Тема 4.2 Реляционные базы данных. Языки запросов реляционных СУБД. Декларативный язык SQL	8	2	4	2	Тест + Практическая работа
5	Тема: 4.3 NoSQL хранилища данных	6	2	2	2	Тест + Практическая работа
6	Тема 4.4 Анализ данных при массово параллельной обработке. Обзор GreenPlum	4	2	0	2	Тест + Практическая работа
7	Тема 4.5 Знакомство с СУБД Postgres. Расширения SQL в	8	2	4	2	

	PostgreSQL					
8	Тема 4.6 Введение в СУБД Clickhouse	6	2	2	2	Тест + Практическая работа
9	Тема 4.7 BI-инструменты. Построение дашбордов с Apache Superset.	12	4	6	2	Тест + Практическая работа
10	Раздел 5. Машинное обучение и оптимизация	30	8	14	8	
11	Тема 5.1 Математические основы машинного обучения	8	2	4	2	Тест + Практическая работа
12	Тема 5.2 Обзор методов анализа больших данных. Выбор и обработка данных для машинного обучения	8	2	4	2	Тест + Практическая работа
13	Тема 5.3 Построение моделей машинного обучения	8	2	4	2	Тест + Практическая работа
14	Тема 5.4 Инструменты анализа данных и машинного обучения (Rapid Miner)	6	2	2	2	Тест + Практическая работа
15	Раздел 6. Нейронные сети	32	8	16	8	
16	Тема 6.1 Введение в нейронные сети. Глубокое обучение	8	2	4	2	Тест + Практическая работа
17	Тема 6.2 Архитектура нейронных сетей. Сверточные и рекуррентные сети	8	2	4	2	Тест + Практическая работа
18	Тема 6.3 Обучение нейронных сетей с помощью TensorFlow и Keras	8	2	4	2	Тест + Практическая работа
19	Тема 6.4 Анализ естественного языка с помощью методов искусственного интеллекта	8	2	4	2	Тест + Практическая работа
20	Раздел 7. Продвижение продукта. Бизнес-метрики	12	4	4	4	
21	Тема 7.1. Введение в продуктовую аналитику	6	2	2	2	Тест + практическая работа
22	Тема 7.2. Ключевые метрики роста	6	2	2	2	Тест +

	продукта					практическая работа
23	Промежуточная аттестация	4				Решение кейса

Содержание модуля 2

Модуль 2. Профилирующий		
Раздел 4. Технологии хранения и обработки больших данных		
Тема 4.1 Введение в теорию БД. Моделирование данных. Основы языка SQL	Лекция (4 ч)	Введение в теорию баз данных. Разновидности баз данных: реляционные, графовые, колоночные сетевые и т. д. Уровни представления данных (ODS, DDL, семантический слой, модель данных). Моделирование данных. Введение в язык SQL. Основы операторов SQL. Команды выбора. Агрегированные команды. Группировка по командам. Понятие СУБД
	Практическое занятие (4 ч)	Разработайте базу данных, предназначенную для хранения логов веб-сервера. Лог включает в себя следующие поля: адрес ресурса (URL), IP-адрес пользовательского компьютера, отметка времени начала просмотра ресурса, длительность просмотра ресурса.
	Самостоятельная работа (2 ч)	Изучить материал: Совместное использование базы данных. Безопасность данных. Тест 4.1: Введение в теорию БД
Тема 4.2 Реляционные базы данных. Языки запросов реляционных СУБД. Декларативный язык SQL	Лекция (2 ч)	Реляционные базы данных для хранения структурированных данных. Сравнение OLAP и OLTP систем. Создание таблиц нормализованных данных. Внедрение денормализованных схем. Погружение в язык SQL. Операторы, оконные функции в языке SQL. Порядок вызова операторов в запросе select
	Практическое занятие (4 ч)	Выполнение множества запросов на закрепление базовых операторов языка SQL на примере PostgreSQL  sql_first_lesson.sql
	Самостоятельная работа (2 ч)	Изучить дополнительные возможности языка SQL: триггеры, хранимые процедуры, функции, индексы и т. д. Алгоритмическая сложность вызова разных типов операторов

		join Тест 4.2: Реляционные базы данных
Тема: 4.3 NoSQL хранилища данных	Лекция (2 ч)	Хранилища данных NoSQL, назначение и особенности. Отличия от реляционных баз данных. NoSQL в BigData. Использование базы данных NoSQL: MongoDB
	Практическое занятие (2 ч)	Разработайте запросы для загрузки данных логов веб-сервера в формате JSON в СУБД MongoDB.
	Самостоятельная работа (2 ч)	Изучить материал: Документно-ориентированная моделью данных MongoDB Тест 4.3: NoSQL хранилища данных
Тема 4.4 Анализ данных при массово параллельная обработка. Обзор GreenPlum	Лекция (2 ч)	Массово параллельные базы данных в Big Data. Что такое MPP-СУБД и как это работает. MPP-СУБД для хранения и аналитики Big Data на примере GreenPlum
	Самостоятельная работа (2 ч)	Изучить материал: Особенности организации СУБД в MPP-системе Тест 4.4: массово параллельная обработка и анализ данных
Тема 4.5 Знакомство с СУБД Postgres. Расширения SQL в PostgreSQL	Лекция (2 ч)	Подробное знакомство с СУБД Postgres. Введение понятий: ограничения, первичный ключ, внешний ключ. Основные типы данных в СУБД Postgres. Ключевые особенности PSQL
	Практическое занятие (4 ч)	Создание таблиц в СУБД Postgres посредством Dbeaver. Таблицы должны быть связаны внешними ключами, при этом сохраняя ссылочную целостность. На данные в таблицах должны быть наложены ограничения. Наполнение таблиц данными с учетом типов данных. Отработка запросов на языке PSQL на созданных таблицах  sql_second.sql
	Самостоятельная работа (2 ч)	Провести сравнение: Postgres с СУБД MS SQL, OracleDB, MySQL. Тест 4.5: Знакомство с СУБД Postgres
Тема 4.6 Введение в СУБД Clickhouse	Лекция (2 ч)	Введение в СУБД Clickhouse. Особенности колоночного формата хранения данных. Ключевые отличия от СУБД Postgres. Преимущества и недостатки
	Практическое занятие (2 ч)	Подключение к СУБД Clickhouse. Создание первых таблиц. Работа с данными, отработка запросов. Аналитические функции в clickhouse. Знакомство с tabix

		 clickhouse_lesson.sql
	Самостоятельная работа (2 ч)	Изучить материал: Возможности миграции между PostgreSQL и Clickhouse. Тест 4.6: Введение в СУБД Clickhouse
Тема 4.7 BI инструменты. Построение дашбордов с Apache Superset	Лекция (4 ч)	Введение в принципы построения дашбордов. Обзор различных BI-инструментов. Какие ключевые задачи решает визуализация с помощью BI-систем. Построение дашбордов. Введение в Apache Superset. Ключевые преимущества и недостатки
	Практическое занятие (6 ч)	Построение первого дашборда с помощью Apache Superset. Подключение к СУБД Postgres. Проработка стиля и оформления дашборда
	Самостоятельная работа (2 ч)	Изучить материал: Возможностей BI-системы Apache Superset. Изучение альтернативных BI-систем. Тест 4.7: BI-инструменты
Раздел 5. Машинное обучение и оптимизация		
Тема 5.1 Математические основы машинного обучения	Лекция (2 ч)	Для чего нужна линейная алгебра в МО. Матрицы и работа матриц в классических моделях. Математическое моделирование. Методы и модели классификации: логистическая регрессия, деревья решений. Методы и модели регрессии: линейная регрессия, деревья решений. Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма. Статистика и оптимизация в машинном обучении
	Практическое занятие (4 ч)	Создание собственной модели линейной регрессии, сравнение работы с готовым решением в Scikit-learn
	Самостоятельная работа (2 ч)	Изучить материал: Генетические алгоритмы, эволюционное программирование Тест 5.1: Математические основы машинного обучения
Тема 5.2 Обзор методов анализа больших данных. Выбор и обработка данных для машинного обучения	Лекция (2 ч)	Построение моделей машинного обучения. Типы задач машинного обучения. Машинное обучение на больших данных. Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация. Методы и модели классификации и регрессии: градиентный бустинг. Метрики качества моделей. Ассоциативные правила. Бизнес-сценарии и

		краткий справочник по машинному обучению. Сравнение алгоритмов машинного обучения. Требования к сценарию обработки и анализа данных.
	Практическое занятие (4 ч)	1. Предобработка данных и построение нескольких моделей и выбор лучшей модели на данных. 2. Произвести поиск аномалий в данных, сегментация PCA, уменьшение размерности данных
	Самостоятельная работа (2 ч)	Изучить материал: Подготовка датасета для машинного обучения, базовые способы совершенствования данных Тест 5.2: Обзор методов анализа больших данных
Тема 5.3 Построение моделей машинного обучения	Лекция (2 ч)	Основные современные и популярные алгоритмы моделей машинного обучения — градиентные бустинги на деревьях и нейронные сети. Алгоритмы кластеризации
	Практическое занятие (4 ч)	Построить модель машинного обучения для поиска мест залегания полезных ископаемых В роли признаков здесь выступают сведения, добытые при помощи геологической разведки: наличие на территории местности каких-либо пород (и это будет признаком бинарного типа), их физические и химические свойства (которые раскладываются на ряд количественных и качественных признаков).
	Самостоятельная работа (2 ч)	Изучить тему: Применение алгоритмов кластеризации данных Тест 5.3: Построение моделей машинного обучения
Тема 5.4 Инструменты анализа данных и машинного обучения (Rapid Miner)	Лекция (2 ч)	Интеллектуальный анализа данных с помощью языка программирования R. Манипуляция, анализ и моделирование данных с помощью RapidMiner. Инструменты анализа данных и ML Rapid Miner
	Практическое занятие (2 ч)	Ознакомление с синтаксисом языка R для анализа данных. Обработка данных с помощью библиотеки tidyverse. Статистический анализ данных в R
	Самостоятельная работа (2 ч)	Изучить материал: Применение Rapid Miner для анализа временных рядов и анализа выделенных параметров Тест 5.4: Инструменты анализа данных и машинного обучения
Раздел 6. Нейронные сети		

Тема 6.1 Введение в нейронные сети. Глубокое обучение	Лекция (2 ч)	Введение в глубокое обучение и нейронные сети. Как работает глубокое обучение. Революция в глубоком обучении. Как устроены многослойные нейронные сети. Как обучают нейросети. Алгоритм обратного распространения ошибки для обучения нейронных сетей. Основные архитектуры и задачи нейронных сетей. Производная, оптимизация
	Практическое занятие (4 ч)	Построение собственной небольшой модели линейной регрессии на основе библиотеки PyTorch
	Самостоятельная работа (2 ч)	Изучить материал: Преимущества глубокого обучения в облаке Тест 6.1: Введение в нейронные сети
Тема 6.2 Архитектура нейронных сетей. Сверточные и рекуррентные сети	Лекция (2 ч)	Продвинутое обучение нейронных сетей. Многослойный перцептрон. Архитектура CNN, свёрточные нейронные сети. Разница между моделями и как устроена архитектура. Комбинирование нескольких алгоритмов нейронных сетей
	Практическое занятие (4 ч)	Построение полносвязной нейронной сети для задачи классификации на основе датасета MNIST
	Самостоятельная работа (2 ч)	Изучить материал: Комбинирование нескольких алгоритмов нейронных сетей Тест 6.2: Архитектура нейронных сетей
Тема 6.3 Обучение нейронных сетей с помощью TensorFlow и Keras	Лекция (2 ч)	Искусственная нейронная сеть в TensorFlow. Как обучить нейронную сеть с помощью TensorFlow и Keras. Чем отличается TensorFlow от PyTorch?
	Практическое занятие (4 ч)	Изменить нейронную сеть (ссылка) Вычислить средние и дисперсии по выборке обучающего датасета и применить их для того, чтобы производить более качественную предобработку данных.
	Самостоятельная работа (2 ч)	Изучить материал: Классификация изображений с помощью TensorFlow и Keras Тест 6.3: Обучение нейронных сетей с помощью TensorFlow и Keras
Тема 6.4 Анализ естественного языка с помощью методов искусственного интеллекта. NLP – обработка естественного языка	Лекция (2 ч)	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
	Практическое занятие (4 ч)	Анализ естественного языка с помощью методов искусственного интеллекта

		Воспользоваться TF-IDF или эмбедингами для анализа токсичности текстов
	Самостоятельная работа (2 ч)	Изучить материал: Модели трансформеров и что такое self-attention Тест 6.4: NLP – обработка естественного языка
Раздел 7. Продвижение продукта. Бизнес-метрики		
Тема 7.1 Введение в продуктовую аналитику	Лекция (2 ч)	Ключевые задачи продуктовой аналитики. Как анализ данных может влиять на развитие продукта. Какие знания о продукте важны для анализа
	Практическое занятие (2 ч)	Решение кейса по анализу продукта с разработкой полного цикла от получения данных, хранения, обработки и вывода результатов – в виде схемы, ключевая задача – определить метрики для анализа качества продукта
	Самостоятельная работа (2 ч)	Изучить материал: Особенности продуктовой аналитики. Изучить реальные кейсы как анализ данных помог вывести продукт на новый уровень Тест 7.1: Основам продуктовой аналитики
Тема 7.2 Ключевые метрики роста продукта	Лекция (2 ч)	Обзор ключевых метрик роста продукта: метрики привлечения (CPI, LTV и т. д.), метрики вовлеченности (DAU/WAU/MAU и т. д.) и метрики производительности
	Практическое занятие (2 ч)	Реализация расчет ключевых метрик роста продукта с помощью Python
	Самостоятельная работа (2 ч)	Изучить материал: Когортный анализ: метрики роста против метрик продукта Тест 7.2: Ключевые метрики роста продукта
<i>Промежуточная аттестация</i>	4 ч.	<i>Решение кейса</i>

Содержание самостоятельной работы слушателей:

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

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

Вопросы для самостоятельной работы в рамках подготовки к промежуточной аттестации:

- совместное использование базы данных;
- безопасность данных;
- дополнительные возможности языка SQL: триггеры, хранимые процедуры, функции, индексы и т. д.;
- алгоритмическая сложность вызова разных типов операторов join;
- документно-ориентированная модель данных MongoDB;
- особенности организации СУБД в MPP-системе;
- сравнение Postgres с СУБД MS SQL, OracleDB, MySQL;
- возможности миграции между PostgreSQL и Clickhouse;

- возможности BI-системы Apache Superset;
- альтернативные BI-системы;
- генетические алгоритмы, эволюционное программирование;
- подготовка датасета для машинного обучения;
- базовые способы совершенствования данных;
- применение алгоритмов кластеризации данных;
- применение Rapid Miner для анализа временных рядов и анализа выделенных параметров;
- преимущества глубокого обучения в облаке;
- комбинирование нескольких алгоритмов нейронных сетей;
- классификация изображений с помощью TensorFlow и Keras;
- модели трансформеров;
- что такое self-attention;
- особенности продуктовой аналитики;
- примеры того, как анализ данных помог вывести продукт на новый уровень;
- когортный анализ: метрики роста против метрик продукта.

Учебно-методическое сопровождение

Обучающие материалы представлены в виде видеолекций, текстовых и графических материалов, размещенных на платформе sprint 1Т.

Список литературы

1. Алекс Петров. Распределенные данные. Алгоритмы работы современных систем хранения информации / Пер. с англ. А. Коцюба. СПб.: Питер, 2021 – 336 с.
2. Бизнес переходит на искусственный интеллект. // РБК+. Решения. #1 Искусственный интеллект, 5 декабря 2022. – URL: <https://plus.rbc.ru/news/638ce98f7a8aa9f3126daaa2> (дата обращения 25.03.2023)
3. Болотова Ю. А. Методы и алгоритмы интеллектуальной обработки цифровых изображений / Друки А. А., Спицын В. Г. – Томск: Томский политехнический университет, 2016. – 208 с.
4. Брокман Д. Что мы думаем о машинах, которые думают: Ведущие мировые учёные об искусственном интеллекте / Пер. с англ. М. Исааков. М.: Альпина нон-фикшн, 2017. – 552 с.
5. Бруссард М. Искусственный интеллект: Пределы возможного / Пер. с англ. Е. Арье. М.: Альпина Паблишер, 2020. – 362 с.
6. Бурков А. Машинное обучение без лишних слов. – СПб.: Питер, 2020. – 192 с.
7. Вигерс Карл, Битти Джой. Разработка требований к программному обеспечению, 3-е изд., дополненное / Пер. с англ. М.: Русская редакция, СПб.: БХВ-Петербург, 2014. – 736 с.
8. Деревянко М. Э., Нилова Н. М. Обзор современных информационных систем управления бизнес-процессами. – Краснодар: ФГБОУ ВО «Кубанский государственный аграрный университет им. И. Т. Трубилина», 2021. – 497 с.
9. Джон Д. Келлехер, Брайан Мак-Нейми, Аоифе д'Арси. Основы машинного обучения для аналитического прогнозирования: алгоритмы, рабочие примеры и тематические исследования / Пер. с англ. Д. Ключин. М.: Диалектика (Вильямс), 2019. – 656 с.
10. Каменнова, М. С. Моделирование бизнес-процессов в 2 ч. Часть 2: Учебник и практикум для бакалавриата и магистратуры / Машков И. В., Крохин В. В. – М.: Юрайт, 2019. – 228 с
11. Карпова И. П. Базы данных. Учебное пособие. – СПб.: Питер, 2021. – 240 с.

12. Ковалев С. М., Ковалев В. М. Настольная книга аналитика. Практическое руководство по проектированию бизнес-процессов и организационной структуры: Практическое руководство. – М.: 1С-Паблишинг, 2020. – 360 с.
13. Куницын А., Зуев Б. Технический анализ: Полный курс. – М.: Альпина Паблишер, 2017. – 880 с.
14. Лекторский В. А. Человек и системы искусственного интеллекта / Васильев С. Н., Макаров В. Л., Хабриева, Т.Я. [и др.] – СПб.: Юридический центр, 2022. – 328 с.
15. Маркус Г., Дэвис Э. Искусственный интеллект: Перегрузка. Как создать машинный разум, которому действительно можно доверять / Пер. с англ. В. Скворцов. М.: Альпина Паблишер, 2021. – 300 с.
16. Мартин Ф. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей / Пер. с англ. И. Рузмайкина. СПб.: Питер, 2019. – 416 с.
17. Элбон К. Машинное обучение с использованием Python. Сборник рецептов / Пер. с англ. Front Cover. Элбон Крис. СПб.: БХВ-Петербург, 2020. – 384 с.
18. Нархид Ния, Шапира Гвен, Палино Тодд. Apache Kafka. Поточковая обработка и анализ данных / Пер. с англ. И. Пальти. СПб.: Питер, 2021. – 320 с.
19. О'Коннелл М. Искусственный интеллект и будущее человечества / Пер. с англ. М. Кудряшова. М.: Бомбора, 2019. – 272 с.
20. Пиковер К. Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей / Пер. с англ. А. Ефимова. М.: Синдбад, 2022. – 250 с.
21. Пратик Д. Искусственный интеллект с примерами на Python / Пер. с англ. А. Гузикевич. М.: Диалектика-Вильямс, 2019. – 448 с.
22. Расставить нейросети // РБК+. Инновации #1 Искусственный интеллект, 5 декабря 2022. – URL: <https://plus.rbc.ru/news/638ce67f7a8aa9e27b22f26e> (дата обращения 25.03.2023)
23. Технологии позволяют учитывать специфику каждой отрасли // РБК+. От первого лица. #1 Искусственный интеллект, 5 декабря 2022. – URL: <https://plus.rbc.ru/news/638ced6c7a8aa9e28f7bf148> (дата обращения 25.03.2023)
24. Томас Д. Внедрение искусственного интеллекта в бизнес-практику. Преимущества и сложности / Пер. с англ. З. Мамедьяров. М.: Альпина Паблишер, 2021 – 316 с.
25. Фальк Ким. Рекомендательные системы на практике. Практическое пособие / Пер. с англ. Д. Павлов. М.: ДМК Пресс, 2020. – 448 с.
26. Мартин Фаулер, Прамодкумар Дж. Садаладж NoSQL: новая методология разработки нереляционных баз данных / Пер. с англ. И. Попова. М.: Вильямс, 2013. – 192 с.
27. Франк Шёнталер, Готфрид Фоссен, Андреас Обервайс, Томас Карле. Бизнес-процессы. Языки моделирования, методы, инструменты / Пер. с англ. Г. Исхакова, А. Абдулнагимов. М.: Альпина Паблишер, 2019. – 264 с.
28. Цветков А. А. Теория и практика бизнес-анализа в ИТ. – М.: Директ-Медиа, 2020. – 150 с.
29. Цзэн Мин. Как Alibaba использует искусственный интеллект в бизнесе: Сетевое взаимодействие и анализ данных / Пер. с англ. К. Батыгин. М.: Альпина Паблишер, 2022. – 360 с.
30. Цифровая экономика от теории к практике: как российский бизнес использует искусственный интеллект / Исслед. РАЭК / НИУ ВШЭ при поддержке Microsoft. – [Электронный ресурс]. – 2019. – 66 с. – Режим доступа: <http://raec.ru/upload/files/190715-ii.pdf> (дата обращения 25.03.2023).

Интернет-ресурсы

1. Воронцов, К. В. Машинное обучение: курс лекций // MachineLearning.ru. – [Электронный ресурс]. – 2023. – Режим доступа: [http://www.recognition.su/wiki/index.php?title=Машинное обучение \(курс лекций%2С_К.В.Воронцов\)](http://www.recognition.su/wiki/index.php?title=Машинное_обучение_(курс_лекций%2С_К.В.Воронцов)) (дата обращения 25.03.2023).
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 25.03.2023)
3. Бесплатные материалы по Data Engineering от преподавателей МФТИ // Московский физико-технический институт (национальный исследовательский университет). – URL: <https://fpmi-edu.ru/free-de> (дата обращения 25.03.2023)
4. Курс «Машинное обучение» // Coursera. – URL: <https://ru.coursera.org/specializations/gcp-data-machine-learning#courses> (дата обращения 25.03.2023)
5. Open Machine Learning Course // Mlcourse.ai. – URL: <https://mlcourse.ai/book/index.html> (дата обращения 25.03.2023)
6. Курсы по искусственному интеллекту // Университет 20.35. – URL: <https://ai.2035.university/> (дата обращения 25.03.2023)
7. Сайт Национального проекта «Цифровая экономика». – URL: <https://xn--80aapampemcchfmo7a3c9ehj.xn--p1ai/projects/tsifrovaya-ekonomika> (дата обращения 25.03.2023)
8. Сайт федерального проекта «Искусственный интеллект» Национального проекта «Цифровая экономика». – URL: <https://xn--80aapampemcchfmo7a3c9ehj.xn--p1ai/projects/tsifrovaya-ekonomika/p-iskusstvennyy-intellekt-p> (дата обращения 25.03.2023)
9. Искусственный интеллект // РБК. – URL: <https://trends.rbc.ru/trends/tag/ai> (дата обращения 25.03.2023)
10. Искусственный интеллект // TAdviser. Государство. Бизнес-технологии. – URL: [https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82_\(%D0%98%D0%98,_Artificial_intelligence,_AI\)](https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82_(%D0%98%D0%98,_Artificial_intelligence,_AI)) (дата обращения 25.03.2023)

Описание системы оценки качества освоения модуля.

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации. Описание оценочных средств, форм и методов контроля представлены в разделе «Формы аттестации и оценочные материалы» программы.

6. ФОРМЫ АТТЕСТАЦИИ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

В систему оценки качества освоения программы входят:

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

Для зачисления на программу потенциальному слушателю необходимо пройти диагностику базовых знаний (вступительное испытание).

6.1 Диагностика базовых знаний (вступительное испытание)

Диагностика базовых знаний (вступительное испытание) проводится в тестовой форме.

1. Как просмотреть содержимое папки через командную строку?
 - a) ld
 - b) cd
 - c) ws
 - d) ls**
2. Какая команда выведет директорию, в которой сейчас находится пользователь?
 - a) pd
 - b) pwd**
 - c) wd
 - d) cd
3. Какая команда используется для поиска файлов в Linux?
 - a) find**
 - b) search
 - c) locate
 - d) lookup
4. Какая команда используется для вывода процесса в консоль?
 - a) ps**
 - b) top
 - c) ls
 - d) dir
5. Через какую команду осуществляется переход из каталога в каталог?
 - a) pd
 - b) rmdir
 - c) cd**
 - d) wd
6. Как получить количество элементов в списке в Python?
 - a) list.length()
 - b) list.size()
 - c) len(list)**
 - d) list.count()
7. Какой тип данных в Python используется для хранения ключ-значение пар?
 - a) Списки
 - b) Кортежи
 - c) Словари**
 - d) Множества
8. Как удалить элемент из списка по индексу в Python?
 - a) list.remove(индекс)
 - b) list.delete(индекс)
 - c) list.pop(индекс)**
 - d) list.slice(индекс)

9. Как объединить два списка в Python?
- a) `list.join(list1, list2)`
 - b) `list.split(list1, list2)`
 - c) `list.combine(list1, list2)`
 - d) `list.extend(list2)`**
10. Как создать множество (set) в Python?
- a) {1, 2, 3}**
 - b) [1, 2, 3]
 - c) (1, 2, 3)
 - d) 1, 2, 3
11. Какая команда используется в Python для определения функций?
- a) def**
 - b) fun
 - c) func
 - d) function
12. Как обратиться к последнему элементу в списке `my_list`?
- a) `my_list[-1]`**
 - b) `my_list[:1]`
 - c) `my_list[-2]`
 - d) `my_list[1:]`
13. Что такое SMA в техническом анализе?
- a) Стандартное отклонение
 - b) Простое скользящее среднее**
 - c) Экспоненциальное скользящее среднее
 - d) Рыночный индекс
14. Какое ключевое слово используется для обновления данных в таблице в SQL?
- a) SELECT
 - b) FROM
 - c) UPDATE**
 - d) WHERE
15. Какое ключевое слово используется для удаления данных из таблицы в SQL?
- a) DELETE FROM**
 - b) ADD
 - c) MODIFY
 - d) DROP
16. Какой оператор используется для объединения двух таблиц в SQL?
- a) INSERT
 - b) UPDATE
 - c) JOIN**
 - d) DELETE
17. Какой метод используется для снижения размерности данных?
- a) Обучение на основе деревьев решений
 - b) Понижение размерности методом главных компонент**
 - c) Метод опорных векторов
 - d) Регрессия
18. Что такое нормализация (масштабирование) данных?
- a) Процесс сокращения размерности пространства признаков
 - b) Процесс поиска оптимальных параметров
 - c) Процесс приведения всех признаков к одному масштабу**
 - d) процесс заполнения пропусков
19. Какой тип графиков используется для отображения изменения показателя по времени?

- a) Гистограмма
 - b) Круговая диаграмма
 - c) Линейный график**
 - d) Столбчатая диаграмма
20. Что такое выбросы в данных?
- a) Метрика качества модели
 - b) Значения, которые сильно отличаются от общей выборки**
 - c) Аналитический эксперимент
 - d) Критерий поиска ближайших соседей

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

6.2 Текущий контроль знаний

Текущий контроль знаний осуществляется в процессе 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 65% правильных ответов.

Примеры тестов

Модуль 1. Основной

Раздел 2. Основы статистики

Тема: Введение в математическую статистику. Статистические методы анализа данных

Вопросы:

1. Статистика как наука изучает:
 - a) Единичные явления
 - b) Массовые явления**
 - c) Периодические события
2. Статистика изучает явления и процессы посредством изучения:
 - a) Определенной информации
 - b) Статистических показателей**
 - c) Признаков различных явлений
3. Основные стадии экономико-статистического исследования включают: а) сбор первичных данных, б) статистическую сводку и группировку данных, в) контроль и управление объектами статистического изучения, г) анализ статистических данных:
 - a) а, б, в
 - b) а, в, г
 - c) а, б, г**
 - d) б, в, г
4. Требуется вычислить средний стаж деятельности работников фирмы: 6, 5, 4, 6, 3, 1, 4, 5, 4, 5. Какую формулу Вы примените?
 - a) Средняя арифметическая**
 - b) Средняя арифметическая взвешенная
 - c) Средняя гармоническая
5. По какой формуле производится вычисление средней величины в интервальном ряду?
 - a) Средняя арифметическая взвешенная**
 - b) Средняя гармоническая взвешенная

Модуль 2. Профилирующий

Раздел 4. Технологии хранения и обработки больших данных

Тема: BI-инструменты. Построение дашбордов с Apache Superset.

Вопросы:

1. Какой язык программирования используется для разработки Superset?
 - a) Java
 - b) JavaScript и Python**
 - c) Ruby
 - d) C++
2. Какие типы графиков доступны в Superset?
 - a) Только столбчатые и линейные
 - b) Только круговые и гистограммы
 - c) Столбчатые, линейные, круговые, гистограммы и многие другие**
 - d) Только сложные трехмерные графики
3. Как можно в Superset автоматически обновлять данные?
 - a) Через API источника данных**
 - b) Через фреймворк визуализации данных
 - c) Через SSH-сессию на сервере
 - d) Обновление данных возможно только вручную
4. Что такое dashboard в Superset?
 - a) Специальный тип графика, который можно добавить на страницу
 - b) Набор графиков и таблиц, которые можно объединить в одном месте для их лучшей визуализации**
 - c) Функция для автоматической генерации отчетов
 - d) Пользовательский интерфейс для работы с базами данных
5. Как добавить нового пользователя в Superset?
 - a) Только через командную строку
 - b) Вручную через интерфейс Superset**
 - c) Только через файл конфигурации Superset
 - d) Новых пользователей нельзя добавлять в Superset

Выполнение практических работ оценивается по следующей методике.

Критерии и шкала оценивания:

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

5-6 баллов: решение в целом найдено, но оно не является предпочтительным и/или имеются логические ошибки;

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

9-10 баллов: решение 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».

Примеры практических работ

Модуль 1. Основной

Раздел 3. Инструменты и технологии анализа данных

Тема 3.3 Введение в визуализацию данных. Python-библиотеки Matplotlib, Seaborn, Plotly.

Задача: Визуализация данных с помощью библиотек Python.

Произвести визуализацию датасет (ссылка):

Построить график компаний с самым большим кол-вом вакансий.

Построить график самых популярных названий в вакансиях.

Построить график самых популярных местоположений вакансий по их численности и построить круговую диаграмму.

Построить график популярности по колонке Type of ownership, относительно численности вакансий.

Построить график популярности по секторам работы.

Построить график, используя сектора работы и средний уровень зарплат в них.

Построить гистограмму распределения рейтинга компаний.

Построить график средних зарплат по популярным компаниям.

Построить по описанию вакансий облако тегов, исследовать самые популярные слова.

В графиках использовать основные цвета сайта hh.ru.

Решения практических заданий отправьте в файле формата .ipynb, используя редакторы кода Jupyter / Google Colab.

Модуль 2. Профилирующий

Раздел 4. Технологии хранения и обработки больших данных

Тема: 4.3 NoSQL хранилища данных

Задача: Разработка запросов в СУБД MongoDB

1. Разработайте и протестируйте запросы на выборку данных из созданных коллекций.

2. Разработайте и протестируйте функции MapReduce для анализа посещаемости ресурсов web-сервера.

Указания к заданию:

Разработайте следующие запросы, используя встроенные в СУБД.

MongoDB-средства программирования на основе парадигмы MapReduce:

1) Выдать список URL ресурсов с указанием суммарной длительности посещения каждого ресурса, упорядоченный по убыванию.

2) Выдать список URL ресурсов с указанием суммарного количества посещений каждого ресурса, упорядоченный по убыванию.

3) Выдать список URL ресурсов с указанием количества посещений каждого ресурса в день за заданный период, упорядоченный URL ресурса и убыванию количества посещений.

4) Выдать список IP-адресов с указанием суммарного количества и суммарной длительности посещений ресурсов, упорядоченный по адресу, убыванию количества и убыванию длительности.

6.3 Промежуточная аттестация

Промежуточная аттестация предусмотрена после 1 и 2 модулей. Она заключается в решении практико-ориентированных задач (кейсов).

Критерии 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.

Шкала оценивания:

не менее 9 баллов – «отлично»;

7-8 баллов – «хорошо»;

5-6 баллов – «удовлетворительно»;

0-4 балла – «неудовлетворительно».

Примеры заданий промежуточной аттестации:

Модуль 1.

Задание 1.

Развернуть простое облако на учебной платформе IT.Sprint с одной виртуальной машиной и одним хранилищем SSD. Сделать скриншоты рабочего instance.

Задание 2.

Спроектируете схему данных и постройте аналитическую витрину СУБД: использовать Vertica (Docker) либо PostgreSQL (Docker), либо BigQuery (GCP). Датасет: Захват данных из Divolte, либо любой из GCP Public Datasets.

Definition of Done:

- DDL объектов
- DML шагов преобразований
- Опционально: Тестирование на наличие ошибок в данных

Ссылку на репозиторий с кодом прислать в чат с преподавателем.

Модуль 2.

Задание 1.

Выгрузка данных из внешних систем. Загрузка сырых данных.

Необходимо настроить data pipeline в Apache NiFi, который будет:

- 1.Получать данные, отправляемые по HTTP (по сути, работать как API).
- 2.Анализировать содержимое того, что ему прислали.
- 3.Выполнять преобразования (склейку).
- 4.Сохранять результат в отдельную директорию в соответствии с контентом.

В результате должны получить:

- 1.Работающий pipeline NiFi в контейнерном окружении.
- 2.Примеры curl-команд, приводящих к разным результатам.
- 3.Текстовые данные в соответствующих директориях контейнерной файловой системы.

Задание 2.

Собрать статистику продаж в филиалах супермаркета и их зависимость от площади магазина и ежедневного количества покупателей, используя Apache Spark.

В качестве исходных данных используется датасет

<https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data>

Выполнить обработку и анализ данных:

- 1.Произвести импорт библиотек и загрузить данные из датасета.
- 2.Проверить данные на корректность, наличие дубликатов. Очистить.
- 3.Провести кластерный анализ данных.
- 4.Выполнить предварительную обработку данных в ML.
- 5.Написать рекомендации.

Ссылку на репозиторий с кодом прислать в чат с преподавателем.

6.4 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 «Эффектон»).

Описание практико-ориентированных задач (кейсов)

Задача 1

Компания разработчик ПО хочет узнать, какие функции и особенности их мобильного приложения являются наиболее востребованными среди пользователей. Для этого

необходимо проанализировать данные об использовании приложения и выявить наиболее часто используемые функции и особенности.

Задача:

- 1.Собрать данные об использовании мобильного приложения компании.
- 2.Проанализировать данные, используя соответствующие инструменты и методы анализа данных.
- 3.Выделить наиболее часто используемые функции и особенности мобильного приложения.
- 4.Подготовить отчет, содержащий результаты анализа, рекомендации по улучшению приложения и графическую визуализацию данных.
- 5.Представить отчет руководству компании и обсудить результаты анализа и рекомендации.

Задача 2

Компания по производству строительного материала заметила, что у них существует большое количество бракованных изделий, которые не проходят тестирование качества. Чтобы решить эту проблему, компания хочет проанализировать данные о производстве и качестве изделий и выявить причины брака.

Задача:

- 1.Собрать данные о производстве и качестве изделий компании.
- 2.Проанализировать данные, используя соответствующие инструменты и методы анализа данных.
- 3.Выделить факторы, которые могут приводить к браку продукции, такие как некачественное сырье, ошибки в процессе производства и т. д.
- 4.Подготовить отчет, содержащий результаты анализа, рекомендации по устранению причин брака и графическую визуализацию данных.
- 5.Представить отчет руководству компании и обсудить результаты анализа и рекомендации.

Задача 3

Ретейл компания запустила новый интернет-магазин и хочет проанализировать данные о продажах, чтобы определить, какие товары наиболее популярны у покупателей и какие маркетинговые кампании наиболее эффективны.

Задача:

- 1.Собрать данные о продажах интернет-магазина компании.
- 2.Проанализировать данные, используя соответствующие инструменты и методы анализа данных.
- 3.Выделить наиболее популярные товары и категории товаров, а также наиболее эффективные маркетинговые кампании.
- 4.Подготовить отчет, содержащий результаты анализа, рекомендации по оптимизации ассортимента товаров и маркетинговых кампаний и графическую визуализацию данных.
- 5.Представить отчет руководству компании и обсудить результаты анализа и рекомендации.

Задача 4

Разрабатывается система распознавания образов на изображениях с использованием глубокого обучения. Однако система имеет высокую ошибку на изображениях с необычными объектами, которых не было в обучающем наборе данных. Задача технического аналитика состоит в том, чтобы улучшить производительность системы на необычных объектах, используя различные методы и техники глубокого обучения.

Задача:

- 1.Собрать дополнительные данные об объектах, которых не было в обучающем наборе данных, с помощью поиска в Интернете или использования других источников данных.

2.Подготовить данные, используя методы предобработки данных, такие как обрезка изображений, изменение размера, нормализация, аугментация и т. д.

3.Обучить и настроить модели глубокого обучения на расширенном наборе данных, используя методы, такие как сверточные нейронные сети, рекуррентные нейронные сети, сверточные рекуррентные нейронные сети, трансформеры и т. д.

4.Оценить производительность моделей на тестовом наборе данных, используя метрики, такие как точность, полнота, F-мера, AUC-ROC и т. д.

5.Выбрать лучшую модель и протестировать ее на реальных изображениях с необычными объектами.

6.Подготовить отчет, содержащий результаты анализа, рекомендации по оптимизации системы распознавания образов и графическую визуализацию данных.

Задача 5

Стартап разрабатывает систему чат-ботов для обслуживания клиентов. Однако некоторые клиенты жалуются на то, что боты не всегда понимают их вопросы или дают неверные ответы. Задача технического аналитика состоит в том, чтобы улучшить производительность системы чат-ботов, используя различные методы и техники обработки естественного языка.

Задача:

1.Собрать данные жалоб клиентов из чат-ботов, собранных через различные каналы, такие как электронная почта, телефон и социальные сети.

2.Провести анализ данных, чтобы понять, какие типы вопросов и проблем наиболее часто возникают у клиентов и какие ответы дают боты на эти вопросы.

3.Проанализировать существующие модели обработки естественного языка, такие как рекуррентные нейронные сети, сверточные нейронные сети, трансформеры и т. д., и выбрать наиболее подходящую модель для решения задачи.

4.Подготовить данные для обучения модели, провести предобработку данных, такую как токенизация, стемминг, удаление стоп-слов и т. д.

5.Обучить и настроить модель на расширенном наборе данных, используя методы, такие как сверточные нейронные сети, рекуррентные нейронные сети, сверточные рекуррентные нейронные сети, трансформеры и т. д.

6.Оценить производительность модели на тестовом наборе данных, используя метрики, такие как точность, полнота, F-мера, AUC-ROC и т. д.

7.Выбрать лучшую модель и интегрировать ее в систему чат-ботов стартапа.

8.Провести тестирование системы на реальных вопросах и проблемах клиентов.

9.Подготовить отчет, содержащий результаты анализа, рекомендации по оптимизации системы чат-ботов и графическую визуализацию данных.

Задача 6

Интернет-провайдер столкнулась с проблемой высокого процента жалоб клиентов на низкую скорость интернет-соединения. Компания хочет определить причины низкой скорости и разработать меры для ее улучшения.

Задача:

1.Собрать данные о скорости интернет-соединения, используемых устройств и других факторах, которые могут влиять на скорость интернета.

2.Проанализировать данные для выявления проблем, таких как области с низкой скоростью интернета, наличие узких мест в сети и другие факторы, которые могут влиять на скорость интернета.

3.Разработать модели машинного обучения, чтобы предсказывать, когда и где будут возникать проблемы с низкой скоростью интернета, и предложить меры для их предотвращения.

4.Определить оптимальную конфигурацию сети, которая позволит увеличить скорость интернета и снизить процент жалоб клиентов на низкую скорость.

5. Предложить рекомендации по улучшению инфраструктуры сети, такие как установка новых маршрутизаторов, увеличение емкости сети и т. д., чтобы улучшить скорость интернета.

6. Оценить эффективность предложенных мероприятий, используя метрики, такие как уровень удовлетворенности клиентов, скорость интернета, время ответа и т. д.

7. Определить возможные риски, связанные с предложенными мероприятиями, и разработать стратегии для их снижения.

8. Разработать план внедрения предложенных мероприятий и обучить персонал компании, чтобы он мог реализовать этот план.

9. Провести тестирование системы на реальных данных, чтобы убедиться в ее эффективности и соблюдении всех регулирующих требований.

10. Определить потенциальную выгоду от использования предложенных мероприятий, такую как увеличение количества клиентов и повышение их лояльности

Задача 7

Разрабатывается система управления микроклиматом в зданиях, используя сенсоры и алгоритмы машинного обучения. Однако пользователи жалуются на неудобство использования системы и низкую точность прогнозирования температуры. Задача технического аналитика состоит в том, чтобы улучшить пользовательский опыт и точность системы управления микроклиматом.

Задача:

1. Собрать данные о температуре, влажности, освещенности и других параметрах микроклимата в зданиях, используя различные типы сенсоров и оборудования.

2. Провести анализ данных, чтобы понять, какие параметры наиболее сильно влияют на температуру в здании и как они взаимосвязаны друг с другом.

3. Проанализировать существующие алгоритмы прогнозирования температуры, такие как линейная регрессия, случайный лес, нейронные сети и т. д., и выбрать наиболее подходящий алгоритм для решения задачи.

4. Подготовить данные для обучения модели, провести предобработку данных, такую как масштабирование, уменьшение размерности, фильтрация шума и т. д.

5. Обучить и настроить модель на расширенном наборе данных, используя методы, такие как линейная регрессия, случайный лес, нейронные сети и т. д.

6. Оценить производительность модели на тестовом наборе данных, используя метрики, такие как средняя абсолютная ошибка (MAE), среднеквадратическая ошибка (MSE), коэффициент детерминации (R-squared) и т. д.

7. Интегрировать алгоритм прогнозирования температуры в систему управления микроклиматом.

8. Провести тестирование системы на реальных зданиях, используя различные сценарии микроклимата и параметры настроек системы.

9. Провести пользовательские тестирования, чтобы определить, какие улучшения могут быть внесены в систему, чтобы сделать ее более удобной для пользователей.

Задача 8

Компания разрабатывает систему автоматической обработки естественного языка для классификации текстовых данных, таких как обращения клиентов в службу поддержки и социальные медиа сообщения. Однако на текущий момент система имеет низкую точность распознавания интенции клиентов и выдачи соответствующего ответа. Задача технического аналитика состоит в улучшении точности системы автоматической обработки естественного языка.

Задача:

1. Собрать большой набор данных обращений клиентов в службу поддержки и социальных медиа сообщений, которые нужно классифицировать.

2. Проанализировать данные и определить наиболее часто встречающиеся категории обращений и интенции клиентов.

3. Создать корпус текстовых данных, содержащих предложения, словосочетания и ключевые слова для каждой категории обращений.

4. Подготовить данные для обучения модели, провести предобработку данных, такую как удаление стоп-слов, лемматизация и т. д.

5. Обучить и настроить модель на расширенном наборе данных, используя методы, такие как классификация на основе правил, наивный Байесовский классификатор, решающие деревья и т. д.

6. Оценить производительность модели на тестовом наборе данных, используя метрики, такие как точность, полнота, F1-мера и т. д.

7. Интегрировать алгоритм классификации текстовых данных в систему автоматической обработки естественного языка компании.

8. Провести тестирование системы на реальных данных обращений клиентов в службу поддержки и социальных медиа сообщений, используя различные сценарии классификации и параметры настроек системы.

9. Провести пользовательские тестирования, чтобы определить, какие улучшения могут быть внесены в систему, чтобы сделать ее более точной и эффективной для обработки данных клиентов.

Задача 9

Разрабатывается система мониторинга пациентов с хроническими заболеваниями, такими как диабет, артериальная гипертензия и бронхиальная астма. Система должна обеспечить более эффективный и точный мониторинг состояния пациентов, чтобы предотвратить возможные осложнения и улучшить их качество жизни. Задача технического аналитика состоит в разработке алгоритмов и моделей, которые могут помочь достичь этих целей.

Задача:

1. Собрать большой набор данных о пациентах с хроническими заболеваниями, включая сведения о медицинской истории, симптомах, лекарствах и результаты лабораторных исследований.

2. Проанализировать данные и определить наиболее важные факторы, влияющие на состояние пациентов и вероятность возникновения осложнений.

3. Создать корпус текстовых данных, содержащих предложения, словосочетания и ключевые слова для каждого заболевания.

4. Подготовить данные для обучения модели, провести предобработку данных, такую как удаление выбросов, заполнение пропущенных значений, нормализацию данных и т. д.

5. Обучить и настроить модели на расширенном наборе данных, используя методы машинного обучения, такие как линейные и нелинейные модели, деревья решений, нейронные сети и т. д.

6. Оценить производительность моделей на тестовом наборе данных, используя метрики, такие как точность, полнота, F1-мера и т. д.

7. Интегрировать алгоритмы мониторинга пациентов в систему компании ABC, которая позволит автоматически мониторить состояние пациентов и предупреждать об возможных осложнениях.

8. Провести тестирование системы на реальных данных пациентов с хроническими заболеваниями, используя различные сценарии мониторинга и параметры настроек системы.

9. Провести пользовательские тестирования, чтобы определить, какие улучшения могут быть внесены.

Задача 10

Клиника занимается лечением пациентов с различными заболеваниями. Для того чтобы повысить удовлетворенность пациентов и увеличить доходы, клиника хочет оптимизировать свои процессы и улучшить качество услуг. Задача технического аналитика состоит в разработке алгоритмов и моделей, которые могут помочь клинике достичь этих целей.

Задача:

1.Собрать данные о пациентах, включая сведения о медицинской истории, симптомах, лекарствах, результаты лабораторных исследований, истории посещений и т. д.

2.Проанализировать данные и определить наиболее важные факторы, влияющие на качество услуг и удовлетворенность пациентов, такие как время ожидания, уровень обслуживания, квалификация врачей и т. д.

3.Разработать алгоритмы и модели для прогнозирования уровня удовлетворенности пациентов, используя методы машинного обучения, такие как линейные и нелинейные модели, деревья решений, нейронные сети и т. д.

4.Разработать модели и алгоритмы для оптимизации процессов клиники, такие как расписание врачей, управление запасами, оптимизация времени ожидания пациентов и т. д.

5.Интегрировать алгоритмы и модели в информационную систему клиники, чтобы автоматизировать процессы и повысить качество обслуживания пациентов.

6.Провести тестирование системы на реальных данных клиники, используя различные сценарии и параметры настроек системы.

7.Провести пользовательские тестирования, чтобы определить, какие улучшения могут быть внесены в систему и процессы клиники.

8.Определить потенциальную выгоду от использования алгоритмов и моделей, такую как повышение качества услуг, улучшение удовлетворенности пациентов и увеличение доходов клиники.

Задача 11

Разрабатывается система для автоматизации диагностики заболеваний на основе анализа медицинских изображений. Система использует методы компьютерного зрения и машинного обучения для распознавания и классификации изображений, чтобы помочь врачам быстрее и точнее диагностировать заболевания и предлагать соответствующие лечения.

Задача:

1.Собрать данные о медицинских изображениях различных заболеваний, таких как рак молочной железы, рак легкого, рак кожи и т. д.

2.Проанализировать данные и определить наиболее важные признаки, влияющие на диагностику заболеваний, такие как форма, текстура, размер, цвет и т. д.

3.Разработать алгоритмы и модели для автоматической диагностики заболеваний, используя методы машинного обучения, такие как сверточные нейронные сети, рекуррентные нейронные сети, глубокие сверточные нейронные сети и т. д.

4.Определить точность и надежность системы, используя метрики, такие как чувствительность, специфичность, точность и площадь под ROC-кривой.

5.Оценить производительность системы, используя метрики, такие как скорость работы, потребление ресурсов и т. д.

6.Интегрировать систему в медицинские учреждения и обеспечить необходимую инфраструктуру для ее функционирования, такую как хранение и обработку больших объемов данных, защиту данных и т. д.

7.Обучить врачей и медицинский персонал использовать систему и давать правильные рекомендации на основе ее результатов.

8.Провести тестирование системы на реальных медицинских изображениях, используя различные сценарии и параметры настроек системы.

9.Провести пользовательские тестирования, чтобы определить, какие улучшения могут быть внесены в систему и процессы медицинских учреждений.

10.Определить потенциальную выгоду от использования системы, такую как повышение точности и скорости диагностики

Задача 12

Компания занимается разработкой системы для мониторинга пациентов в реальном времени с помощью носимых устройств и приложений для смартфонов. Система использует алгоритмы машинного обучения для анализа данных, получаемых от датчиков на носимых

устройствах, чтобы предоставлять врачам информацию о состоянии пациентов и помогать им принимать более эффективные решения.

Задача:

1. Разработать систему мониторинга пациентов, включая носимые устройства, приложения для смартфонов и облачную инфраструктуру для хранения и анализа данных.

2. Определить метрики и параметры, которые необходимы для мониторинга пациентов, такие как частота сердечных сокращений, давление, уровень кислорода в крови и т. д.

3. Разработать алгоритмы и модели машинного обучения для анализа данных, получаемых от датчиков на носимых устройствах, чтобы предоставлять врачам информацию о состоянии пациентов и помогать им принимать более эффективные решения.

4. Оценить точность и надежность системы мониторинга пациентов, используя метрики, такие как чувствительность, специфичность и точность.

5. Оценить производительность системы, используя метрики, такие как скорость работы, потребление ресурсов и т. д.

6. Обеспечить защиту данных, включая конфиденциальность и целостность данных пациентов, и соблюдать соответствующие регулирующие требования.

7. Обучить врачей и медицинский персонал использовать систему и давать правильные рекомендации на основе ее результатов.

8. Провести тестирование системы на реальных данных пациентов, используя различные сценарии и параметры настроек системы.

9. Провести пользовательские тестирования, чтобы определить, какие улучшения могут быть внесены в систему и процессы медицинских учреждений.

10. Определить потенциальную выгоду от использования системы, такую как улучшение качества медицинского ухода, снижение

Задача 13

Телекоммуникационная компания, являющаяся поставщиком услуг сотовой связи, хочет оптимизировать свою сеть, чтобы улучшить качество обслуживания своих клиентов. В частности, компания хочет оптимизировать использование частотного спектра для улучшения качества сигнала и снижения задержки в сети.

Задача:

1. Собрать данные о использовании сети в различных локациях, таких как данные о потреблении трафика, качестве сигнала и задержке.

2. Проанализировать данные для выявления проблем в использовании сети, таких как области с низким качеством сигнала или высокой задержкой, и определить причины этих проблем.

3. Разработать модели машинного обучения, чтобы предсказывать, когда и где будут возникать проблемы в использовании сети, и предложить меры для их предотвращения.

4. Определить оптимальное использование частотного спектра для улучшения качества сигнала и снижения задержки в сети.

5. Предложить рекомендации по оптимизации инфраструктуры сети, такие как установка новых вышек связи, улучшение маршрутизации сигналов и т. д., чтобы улучшить качество обслуживания клиентов.

6. Оценить эффективность предложенных мероприятий, используя метрики, такие как уровень удовлетворенности клиентов, качество сигнала, скорость передачи данных и т. д.

7. Определить возможные риски, связанные с предложенными мероприятиями, и разработать стратегии для их снижения.

8. Разработать план внедрения предложенных мероприятий и обучить персонал компании, чтобы он мог реализовать этот план.

9. Провести тестирование системы на реальных данных, чтобы убедиться в ее эффективности и соблюдении всех регулирующих требований.

10. Определить потенциальную выгоду от использования предложенных мероприятий, такую как увеличение количества клиентов и повышение их лояльности, а также увеличение доходов компании.

Задача 14

Компания рассматривает возможность перехода на облачные вычисления в своей ИТ-инфраструктуре. Компания хочет, чтобы технический аналитик провел анализ технологических рисков, которые могут возникнуть в процессе перехода на облачные вычисления.

Задачи:

1. Определить, какие сервисы и приложения будут перенесены на облако.
2. Определить, какие данные будут храниться в облаке и какие будут оставаться в локальной инфраструктуре.
3. Определить, какие изменения в существующей архитектуре системы необходимо внести для поддержки облачных вычислений.
4. Определить потенциальные риски и угрозы безопасности, связанные с переходом на облако, и рассмотреть возможные меры по их снижению.
5. Определить требования к провайдеру облачных вычислений в области безопасности данных и управления рисками.
6. Определить необходимые ресурсы для переноса системы в облако и рассмотреть возможные затраты.
7. Определить необходимость обучения сотрудников компании, которые будут работать с облачной инфраструктурой.

Результаты:

Технический аналитик должен представить доклад, включающий:

- Описание системы, которая будет перенесена на облако.
- Описание облачного провайдера, установленных мер безопасности и управления рисками.
- Анализ технологических рисков и угроз безопасности, а также описать возможные меры по их снижению.
- Оценку затрат на перенос системы в облако.
- План обучения сотрудников.

Задача 15

Вы являетесь техническим аналитиком в крупной технологической компании. Управление компании обратилось к вам с просьбой разработать систему прогнозирования возможных кибератак на их системы и сети. Компания хочет иметь возможность оперативно реагировать на угрозы и предотвращать наиболее серьезные атаки.

Задача:

Ваша задача — разработать систему прогнозирования кибератак на крупные технологические компании. Система должна анализировать данные о предыдущих атаках, информации о новых уязвимостях, интернет-трафике и других факторах, которые могут свидетельствовать о возможных угрозах. Система должна обновляться в реальном времени и включать в себя следующие элементы:

1. Сбор данных: разработка системы для сбора данных о предыдущих атаках, уязвимостях и других факторах, которые могут свидетельствовать о возможных угрозах. Эта система должна быть взаимодействовать с другими компонентами вашей системы прогнозирования.
2. Анализ данных: разработка системы анализа данных, которая будет обрабатывать данные, собранные в вашей системе сбора данных. Эта система должна определять наиболее вероятные угрозы и предоставлять операторам компании информацию о том, какие меры нужно предпринимать.
3. Визуализация данных: разработка системы визуализации данных, которая будет отображать информацию о возможных угрозах на графиках и диаграммах. Это позволит

операторам быстро понимать, какие угрозы наиболее серьезны и какие меры нужно предпринять.

4. Мониторинг: разработка системы мониторинга, которая будет отслеживать сетевую активность и анализировать данные на предмет угроз. Эта система должна быть взаимодействовать с другими компонентами вашей системы прогнозирования и оперативно реагировать на угрозы.

Результат:

В результате вашей работы компания получит систему, которая позволит ей оперативно реагировать на угрозы и предотвращать наиболее серьезные кибератаки. Система будет обновляться в реальном времени и включать в себя следующие элементы: систему сбора данных, систему анализа данных, систему визуализации данных и систему мониторинга.

Задача 16

Клиентский опыт является ключевым фактором в успехе интернет-магазина. Надо искать способы улучшения клиентского опыта, чтобы клиенты были лояльны магазину и совершали покупки.

Мы столкнулись с проблемой, что количество покупателей в магазине снизилось, а процент отказов при оформлении заказа увеличился. Необходимо проанализировать данные о клиентах и найти способы оптимизации клиентского опыта.

Цели:

1. Проанализировать поведение клиентов на сайте.
2. Определить причины отказов при оформлении заказа и предложить меры по оптимизации процесса.
3. Изучить аналоги и предложить инновационные идеи для улучшения клиентского опыта.
4. Разработать план действий для внедрения изменений и оценить их эффективность.

Задачи:

1. Провести анализ данных о клиентах и построить графики для визуализации поведения клиентов на сайте.
2. Изучить отзывы клиентов на сайте и определить основные причины отказов при оформлении заказа.
3. Провести анализ существующих процессов оформления заказа и предложить меры по оптимизации.
4. Изучить аналоги и предложить инновационные идеи для улучшения клиентского опыта.
5. Разработать план действий для внедрения изменений и оценить их эффективность.

Инструменты:

1. Язык программирования Python для анализа данных.
2. Библиотеки Pandas, NumPy, Matplotlib для работы с данными и построения графиков.
3. SQL для работы с базами данных.
4. Технология A/B-тестирования для оценки эффективности изменений.

Ожидаемые результаты:

1. Построение графиков для визуализации поведения клиентов на сайте.
2. Определение основных причин отказов при оформлении заказа.
3. Предложение мер по оптимизации процесса оформления заказа.
4. Предложение инновационных идей для улучшения клиентского опыта.
5. Разработка плана действий для внедрения изменений и оценка их эффективности.

Задача 17

Университетская библиотека имеет множество книг и журналов по различным тематикам. Необходимо провести анализ данных, чтобы создать систему рекомендаций для пользователей библиотеки. Рекомендации должны быть основаны на предпочтениях пользователей, а также на их истории использования материалов библиотеки.

Ожидаемый результат:

Результатом работы будут являться:

- Набор данных, содержащий информацию о пользователях, их предпочтениях и истории использования материалов библиотеки.
- Система рекомендаций, основанная на данных, с использованием алгоритмов машинного обучения.

Задачи:

- 1.Собрать данные о пользователях библиотеки, их предпочтениях и истории использования материалов.
- 2.Очистить и подготовить данные для анализа.
- 3.Провести анализ данных и выделить основные характеристики пользователей и материалов, а также выявить скрытые зависимости.
- 4.Разработать систему рекомендаций с использованием алгоритмов машинного обучения.
- 5.Протестировать систему рекомендаций на новых данных и сравнить ее результаты с уже существующими рекомендациями.

Инструменты и технологии:

- Язык программирования Python с использованием библиотек Pandas, NumPy, Scikit-Learn и других.
- СУБД для хранения и обработки данных.
- Инструменты визуализации данных, такие как Matplotlib, Seaborn.
- Рекомендации:
- При сборе данных нужно убедиться, что они анонимизированы и не нарушают конфиденциальность пользователей.
- При разработке системы рекомендаций необходимо учитывать, что она должна быть простой и удобной для использования пользователями, а также устойчивой к изменяющимся предпочтениям пользователей.
- Необходимо провести тестирование системы рекомендаций на реальных данных, чтобы убедиться в ее эффективности и точности.

Задача 18

У вас есть задача как технического аналитика по оценке и улучшению безопасности IoT устройств. Компания занимается производством и продажей интеллектуальных устройств для дома, таких как термостаты, системы безопасности, освещение и т. д. Компания хочет улучшить безопасность своих устройств, чтобы предотвратить взлом и кражу данных клиентов. Одной из основных проблем является то, что устройства продаются с дефолтными паролями, которые многие пользователи не меняют.

Задача:

Вам нужно предложить план действий, чтобы улучшить безопасность устройств компании и наладить процесс обеспечения безопасности устройств на всех этапах жизненного цикла продукта.

1.Изучение текущих методов безопасности

Вам нужно изучить текущие методы безопасности, которые используются в устройствах компании. Это включает в себя безопасность операционной системы, шифрование данных, авторизацию пользователей и т. д. Проведите анализ рисков и определите, какие уязвимости могут быть использованы злоумышленниками для взлома устройств.

2.Разработка плана улучшения безопасности

На основе результатов анализа рисков разработайте план улучшения безопасности устройств. Ваш план должен включать:

- Изменение паролей по умолчанию на сильные и уникальные пароли для каждого устройства
- Включение двухфакторной аутентификации для доступа к устройствам
- Улучшение процедур шифрования данных
- Разработка процедур безопасной установки и настройки устройств

- Обучение пользователей безопасному использованию устройств
- Установка механизмов обнаружения взломов

3. Реализация и тестирование плана

После того как вы разработали план улучшения безопасности, необходимо реализовать и протестировать его.

После реализации плана необходимо провести тестирование безопасности устройств, чтобы убедиться, что новые меры действительно работают и устройства защищены от взлома.

4. Поддержание безопасности устройств

Наконец, необходимо разработать процедуры поддержания безопасности устройств на всех этапах жизненного цикла продукта. Это включает в себя:

- Регулярное обновление прошивок
- Отслеживание уязвимостей и устранение их повторного появления
- Обучение пользователей безопасному использованию устройств
- Систематическое тестирование безопасности устройств

Итоги

После выполнения этой задачи вы должны предоставить компании план действий по улучшению безопасности устройств, а также провести реализацию, тестирование и поддержание мер по обеспечению безопасности устройств. В результате компания сможет предоставить более защищенные устройства для своих клиентов и предотвратить потенциальные проблемы безопасности в будущем.

Задача 19

Мы являемся частью медицинской организации, которая отвечает за создание системы мониторинга заболеваемости населения с целью определения наиболее распространенных заболеваний и проведения исследований по их лечению. Наша организация занимается сбором и обработкой данных о заболеваемости населения на территории города.

Задачи:

1.Собрать данные о заболеваемости населения за последние 10 лет.

Данные должны содержать информацию о каждом инциденте заболевания, такую как вирус или болезнь, возраст пациента, пол, дата и место обращения за медицинской помощью.

2.Установить приоритеты по наиболее распространенным заболеваниям.

Определить наиболее часто встречающиеся заболевания, исходя из имеющихся данных.

Сделать выводы о наиболее уязвимых группах населения – дети, пожилые люди, женщины, мужчины и др.

3.Создать систему мониторинга заболеваемости.

Разработать программу, которая будет автоматически собирать данные о заболеваемости населения на территории города.

Информация должна быть представлена в удобном для анализа виде, с графиками и таблицами.

4.Создание отчетов и результатов анализа данных.

Подготовить периодические отчеты на основе собранных данных, которые будут использоваться для проведения дальнейших исследований.

Подготовить аналитический отчет с рекомендациями о том, какие виды медицинских исследований необходимо провести для улучшения ситуации в медицинской организации.

Задача 20

Команда разработки приложения для анализа данных столкнулась с проблемой низкой производительности при обработке больших объемов данных. В настоящий момент процессы обработки не оптимизированы, а также отсутствует система отслеживания процессов обработки данных.

Задача:

Ваша задача в качестве технического аналитика состоит в том, чтобы создать систему отслеживания процессов обработки данных и оптимизировать их производительность.

Требования:

- Система отслеживания должна предоставлять информацию о процессах обработки данных, таких как объем данных, время выполнения, использование ресурсов и т. д.
- Система должна предупреждать об ошибках и проблемах производительности в режиме реального времени.
- Система должна сохранять исторические данные для анализа производительности в будущем.
- Оптимизация производительности должна включать в себя оптимизацию алгоритмов обработки данных, оптимизацию запросов к базе данных и оптимизацию использования ресурсов.

Решение:

Для решения поставленной задачи будут использоваться следующие шаги:

1. Анализ процессов обработки данных — проведение анализа текущих процессов для определения узких мест и проблем производительности.

2. Создание системы отслеживания процессов обработки данных — реализация мониторинга процессов обработки данных и оповещение об ошибках и проблемах производительности.

3. Оптимизация алгоритмов обработки данных — определение оптимальных алгоритмов обработки данных для ускорения процессов обработки.

4. Оптимизация запросов к базе данных — проведение анализа запросов к базе данных и оптимизация их для ускорения процессов обработки данных.

5. Оптимизация использования ресурсов — проведение анализа использования ресурсов и оптимизация для улучшения производительности.

Ожидаемые результаты:

Создание системы отслеживания процессов обработки данных.

Увеличение производительности приложения для анализа данных.

Сокращение времени обработки данных.

Уменьшение использования ресурсов приложения.

Результатом проекта будет отчет в виде страницы на github

Задача 21

Технологическая компания, занимающаяся разработкой программного обеспечения, запустила новый проект в виде облачного сервиса для управления проектами. Наша задача в качестве технического аналитика — провести исследование рынка и создать план маркетинга для продвижения продукта.

Шаги проекта:

1. Изучение конкурентов

Составить список основных конкурентов на рынке облачных сервисов по управлению проектами

Изучить основные преимущества и недостатки конкурентов

Определить ценовой диапазон на рынке

2. Исследование целевой аудитории

Определить целевую аудиторию продукта

Изучить потребности и интересы целевой аудитории

Определить, какие каналы коммуникации наиболее эффективны

3. Создание уникального предложения

Определить уникальные преимущества и особенности продукта

Создать уникальное предложение, которое будет привлекать потенциальных клиентов

4. Разработка маркетинговой стратегии

Определить цели маркетинговой кампании

Определить бюджет маркетинговой кампании

Разработать маркетинговую стратегию для продвижения продукта

5. Продвижение продукта

Организовать рекламную кампанию

Разработать контентную стратегию

Использовать социальные сети для продвижения продукта

6. Оценка результатов

Оценить результаты маркетинговой кампании

Сравнить результаты с поставленными целями

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

Знать:

- базовые понятия искусственного интеллекта и BigData;
- отличительные особенности машинного обучения, нейронных сетей;
- основы программирования на языке Python;
- ключевые компоненты и парадигмы Hadoop;
- основы теории алгоритмов и структур данных;
- подходы и принципы работы с базами данных NoSQL;
- приемы и принципы управления СУБД на основе языка запросов SQL;
- базовые модели и уровни представления данных;
- основные функции и инструменты процесса ETL;
- преимущества организации MPP- СУБД для хранения и аналитической обработки больших данных;
- особенности и архитектура СУБД Postgres;
- принципы действия, преимущества и область применения колоночного формата хранения данных;
- правила и принципы создания отчетов в контексте визуализации данных;
- фундаментальные основы теории вероятности;
- основные положения комбинаторики;
- суть метода и базовые инструменты А/В-тестирования;
- подходы, принципы и инструменты продуктовой аналитики;
- аналитический инструментарий и перспективные методики технологий больших данных.
- классические модели машинного обучения и популярные архитектуры нейронных сетей.

Уметь:

- анализировать широкий спектр информации и данных с целью повышения эффективности конкретных ресурсов/процессов или внедрения решений по автоматизации;
- осуществлять решение прикладных задач с использованием SQL базы данных;
- разрабатывать алгоритмы прикладных решений на основе ИИ с использованием языка программирования и библиотек Python;
- применять современные методы и инструменты поиска и структурирования данных;
- применять технику визуализации для интеллектуального анализа данных;
- применять принципы интеллектуального анализа данных;
- применять графовые модели анализа данных;
- выполнять исследования и расчеты показателей проекта на основе методов имитационного моделирования вероятностных распределений с использованием языка программирования Python;

- проводить исследования поискового и аналитического характера с помощью инструментов технического анализа на основе методов и техник анализа больших данных;
- создавать и внедрять математические и информационные модели в зависимости от целей аналитических исследований;
- проектировать технические решения проектов в сфере информационных технологий и обеспечивать их поддержку на протяжении всего жизненного цикла.

Владеть:

- навыками решения ситуационных аналитических задач с использованием алгоритмов искусственного интеллекта и инструментов визуализации для создания интерактивных дашбордов;
- статистическими методами анализа данных;
- навыками разведочного, описательного и исследовательского анализа данных с использованием языка Python и его инструментальных средств для расчета показателей;
- методами построения моделей машинного обучения и нейронных сетей;
- математическими методами анализа и обработки технологической информации;
- навыками разработки таблиц в СУБД Postgres и управления ими с помощью Dbeaver;
- навыками создания полносвязной нейронной сети для решения задач классификации на основе любых наборов данных;
- навыками использования фреймворков Scikit-learn, Scipy для обучения моделей машинного обучения;
- навыками использования фреймворков PyTorch, TensorFlow и Keras для обучения нейронных сетей;
- навыками анализа финансовых и продуктовых метрик и поиска «слабых» мест с помощью Python.

Критерии оценивания, шкала оценивания

Критерии оценивания:

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

5-6 баллов: решение кейса в целом найдено, но оно не является предпочтительным 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.

Шкала оценивания:

не менее 9 баллов – «отлично»;

7-8 баллов – «хорошо»;

5-6 баллов – «удовлетворительно»;

0-4 балла – «неудовлетворительно».

7. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

7.1 Специфика организационных действий и педагогических условий

Обучение реализуется с применением технологий электронного обучения и дистанционных образовательных технологий.

Для достижения планируемых результатов обучение строится с использованием следующих:

методов: метод решения ситуационных задач, проектный метод, модульное обучение, проблемное обучение, знаково-контекстное обучение;

форм: интерактивные лекции, практические занятия, самостоятельная работа.

7.2 Кадровое обеспечение программы (преподавательский состав)

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

Представители образовательной организации высшего образования и научной организации имеют высшее образование, ученую степень кандидата наук либо PhD, стаж научно-педагогической работы более трех лет или выше, 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 и в журналах Web of Science (WOS) и Scopus.

№ 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	Балыкина Анна Михайловна	АНО ВО «Российский новый университет», доцент общеуниверситетской кафедры основ математики и информатики, канд. психол. наук	
2	Белоглазов Александр Анатольевич	ФГБОУ ВО «Московский государственный гуманитарно-экономический университет», доцент кафедры информационных технологий и прикладной математики, канд. техн. наук	
3	Золотарев Олег Васильевич	АНО ВО «Российский новый университет», заведующий кафедрой «Информационные системы в экономике и управлении» Института информационных систем и инженерно-компьютерных технологий доцент, канд. техн. наук	http://professor.rosnou.ru/?q=taxonomy/term/114

4	Лабунец Леонид Витальевич	АНО ВО «Российский новый университет», профессор кафедры «Информационные системы в экономике и управлении» Института информационных систем и инженерно-компьютерных технологий, докт. техн. наук	
5	Михалёва Елизавета Вячеславовна	АНО ВО «Российский новый университет», заместитель исполнительного директора Института информационных систем и инженерно-компьютерных технологий	
6	Палкин Евгений Алексеевич	АНО ВО «Российский новый университет», проректор по научной работе и заместитель председателя Учёного совета, канд. физ.-мат. наук, профессор, почетный работник высшего профессионального образования Российской Федерации, лауреат Государственной премии СССР	https://clck.ru/33uZcB

7.3 Учебно-методическое обеспечение и информационное сопровождение

Учебно-методические материалы
Методические разработки, материалы курса, учебная литература
<p>Обучающие материалы представлены в виде видеолекций, текстовых и графических материалов, размещенных на платформе sprint.1Т.</p> <p style="text-align: center;"><i>Список учебной и учебно-методической литературы</i></p> <p style="text-align: center;">Иностранные переводные издания</p> <ol style="list-style-type: none"> 1. Брокман Д. Что мы думаем о машинах, которые думают: Ведущие мировые учёные об искусственном интеллекте / Пер. с англ. М. Исааков. М.: Альпина нон-фикшн, 2017. 2. Бруссард М. Искусственный интеллект: Пределы возможного / Пер. с англ. Е. Арье. М.: Альпина Паблишер, 2020. 3. Вигерс Карл, Битти Джой. Разработка требований к программному обеспечению, 3-е изд., дополненное / Пер. с англ. М.: Русская редакция, СПб.: БХВ-Петербург, 2014. 4. Дэви Силен, Арно Мейсман, Мохамед Али. Основы Data Science и Big Data. Python и наука о данных / Пер. с англ. Е. Матвеев. СПб.: Питер, 2018. 5. Джон Д. Келлехер, Брайан Мак-Нейми, Аоифе д'Арси. Основы машинного обучения для аналитического прогнозирования: алгоритмы, рабочие примеры и тематические исследования / Пер. с англ. Д. Ключин. М.: Диалектика (Вильямс), 2019. 6. Конвински Энди, Венделл Патрик, Захария Матей, Карау Холден. Изучаем Spark. Молниеносный анализ данных / Пер. с англ. М.: ДМК Пресс, 2015. 7. Элбон К. Машинное обучение с использованием Python. Сборник рецептов / Пер. с англ. Front Cover. Элбон Крис. СПб.: БХВ-Петербург, 2020.

8. Томас Д. Внедрение искусственного интеллекта в бизнес-практику. Преимущества и сложности / Пер. с англ. З. Мамедьяров. М.: Альпина Паблишер, 2021
9. Пратик Д. Искусственный интеллект с примерами на Python / Пер. с англ. А. Гузикович. М.: Диалектика-Вильямс, 2019.
10. Франк Шёнталер, Готфрид Фоссен, Андреас Обервайс, Томас Карле. Бизнес-процессы. Языки моделирования, методы, инструменты / Пер. с англ. Г. Исхакова, А. Абдулнагимов. М.: Альпина Паблишер, 2019.
11. Фальк Ким. Рекомендательные системы на практике. Практическое пособие / Пер. с англ. Д. Павлов. М.: ДМК Пресс, 2020.
12. Мартин Фаулер, Прамодкумар Дж. Садаладж NoSQL: новая методология разработки нереляционных баз данных / Пер. с англ. И. Попова. М.: Вильямс, 2013.
13. О'Коннелл М. Искусственный интеллект и будущее человечества / Пер. с англ. М. Кудряшова. М.: Бомбора, 2019.
14. Маркус Г., Дэвис Э. Искусственный интеллект: Перегрузка. Как создать машинный разум, которому действительно можно доверять / Пер. с англ. В. Скворцов. М.: Альпина Паблишер, 2021.
15. Мартин Ф. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей / Пер. с англ. И. Рузмайкина. СПб.: Питер, 2019.
16. Нархид Ния, Шапира Гвен, Палино Тодд. Apache Kafka. Поточковая обработка и анализ данных / Пер. с англ. И. Пальти. СПб.: Питер, 2021.
17. Алекс Петров. Распределенные данные. Алгоритмы работы современных систем хранения информации / Пер. с англ. А. Коцюба. СПб.: Питер, 2021
18. ПикOVER К. Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей / Пер. с англ. А. Ефимова. М.: Синдбад, 2022.
19. Цзэн Мин. Как Alibaba использует искусственный интеллект в бизнесе: Сетевое взаимодействие и анализ данных / Пер. с англ. К. Батыгин. М.: Альпина Паблишер, 2022.

Российские издания

1. Бурков А. Машинное обучение без лишних слов. – СПб.: Питер, 2020. – 192 с.
2. Иванова В., Перерва А. Путь аналитика. Практическое руководство IT-специалиста, 2-е изд. – СПб.: Питер, 2015. – 304 с.
3. Болотова Ю. А. Методы и алгоритмы интеллектуальной обработки цифровых изображений / Друки А. А., Спицын В. Г. – Томск: Томский политехнический университет, 2016. – 208 с.
4. Миронов В. Профессия «бизнес-аналитик» Краткое пособие для начинающих. – М.: Альпина Паблишер, 2021. – 192 с.
5. Деревянко М. Э., Нилова Н. М. Обзор современных информационных систем управления бизнес-процессами. – Краснодар: ФГБОУ ВО «Кубанский государственный аграрный университет им. И. Т. Трубилина», 2021. – 497 с.
6. Еременко К. Работа с данными в любой сфере: как выйти на новый уровень, используя аналитику. – М.: Альпина Паблишер, 2019. – 304 с.
7. Каменнова, М. С. Моделирование бизнес-процессов в 2 ч. Часть 2: Учебник и практикум для бакалавриата и магистратуры / Машков И. В., Крохин В. В. – М.: Юрайт, 2019. – 228 с.
8. Карпова И. П. Базы данных. Учебное пособие. – СПб.: Питер, 2021. – 240 с.
9. Ковалев С. М., Ковалев В. М. Настольная книга аналитика. Практическое руководство по проектированию бизнес-процессов и организационной структуры: Практическое руководство. – М.: 1С-Паблишинг, 2020. – 360 с.
10. Косников С. Н. Математические методы в экономике: учебное пособие для вузов, 2-е изд., испр. и доп. – М.: Юрайт, 2019. – 172 с.

11. Кулик Б.А. Логика и математика: просто о сложных методах логического анализа. – СПб.: Политехника, 2020. – 141 с.
12. Куницын А., Зуев Б. Технический анализ: Полный курс. – М.: Альпина Паблишер, 2017. – 880 с.
13. Лекторский В. А. Человек и системы искусственного интеллекта / Васильев С. Н., Макаров В. Л., Хабриева, Т.Я. [и др.] – СПб.: Юридический центр, 2022. – 328 с.
14. Лукьянова Н. Ю., Галицкая Е. Г. Аналитические методы исследований в цифровой экономике. – Казань: Бук, 2019. – 232 с.
15. Поляков В. М., Агаларов З. С. Методы оптимизации: Учебное пособие. – М.: Дашков и К., 2022. – 86 с.
16. Цветков А. А. Теория и практика бизнес-анализа в ИТ. – М.: Директ-Медиа, 2020. – 150 с.

Учебно-методические и информационные материалы

1. 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)Понкин, И.В.; Лаптева А.И. Методология научных исследований и прикладной аналитики: учебник – 2 изд., доп. и перераб. – 2021.
2. Понкин И.В., Лаптева А.И. Методология научных исследований и прикладной аналитики: Учебник. Издание 2-е, дополн. и перераб. / Консорциум «Аналитика. Право. Цифра». – М.: Буки Веди, 2021. – 567 с. (Серия: «Методология и онтология исследований»)
3. Рафалович В. Data mining, или Интеллектуальный анализ данных для занятых. Практический курс / В. Рафалович – Litres, 2022. – 164 с.
4. Системный анализ : учебник и практикум для вузов / В. В. Кузнецов [и др.] ; под общей редакцией В. В. Кузнецова. — Москва: Издательство Юрайт, 2023. — 270 с. — (Высшее образование). — ISBN 978-5-9916-8591-7.Бизнес переходит на искусственный интеллект. / РБК+. Решения. #1 Искусственный интеллект, 5 декабря 2022. – Режим доступа: <https://plus.rbc.ru/news/638ce98f7a8aa9f3126daaa2>
5. Цифровая экономика от теории к практике: как российский бизнес использует искусственный интеллект / Исслед. РАЭК / НИУ ВШЭ при поддержке Microsoft. – [Электронный ресурс]. – 2019. – 66 с. – Режим доступа: <http://raec.ru/upload/files/190715-ii.pdf> (дата обращения 25.03.2023).
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 25.03.2023).
7. Расставить нейросети / РБК+. Инновации. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс]. – Режим доступа: <https://plus.rbc.ru/news/638ce67f7a8aa9e27b22f26e> (дата обращения 25.03.2023).
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 25.03.2023).

Статьи

1. Акулин Е. В. Специфика и особенности задач системного анализа // Актуальные проблемы теории и практики развития научных. – 2022. – С. 20.
2. Алфимов В. А. Использование R/S-анализа и фрактальной теории при анализе финансовых временных рядов // Современные наука и образование: достижения и перспективы развития. – 2021. – С. 8–13.

3. Антипова К. Г. Способы определения больших данных: Российский и зарубежный опыт // Юридические исследования. – 2021. – № 9. – С. 143–157. – URL: https://nbpublish.com/library_read_article.php?id=36591
4. Баева В. Р., Дроздов А. Ю. ETL: Актуальность и применение, преимущества и недостатки ETL инструментов // Вестник науки. – 2019. – №5 (14). – URL: <https://cyberleninka.ru/article/n/etl-aktualnost-i-primenenie-preimuschestva-i-nedostatki-etl-instrumentov>
5. Будасова В. А. Методы технического анализа рынка // Цифровая экономика — инструмент и среда общественного развития. – 2021. – С. 18–21.
6. Городнова Н. В. Применение искусственного интеллекта в бизнес-сфере: современное состояние и перспективы // Вопросы инновационной экономики. – 2021. – Том 11. – № 4. – С. 1473–1492.
7. Доржиева В. В. Цифровизация промышленности: роль искусственного интеллекта и возможности для России // Вопросы инновационной экономики. – 2022. – Т. 12. № 4. – С. 2383–2394.
8. Еременко К. Работа с данными в любой сфере. Как выйти на новый уровень, используя аналитику. – 2018. – С. 20–58.
9. Звягин Л. С. Использование прикладного системного анализа как инструмента моделирования для управления бизнесом // Хроноэкономика. – 2019. – №. 7 (20). – С. 26–31.
10. ОСНОВЫ СИСТЕМНОГО АНАЛИЗА // Столыпинский вестник. – 2022. №4. – URL: <https://cyberleninka.ru/article/n/osnovy-sistemnogo-analiza>
11. Кондуров И. В., Тушев А. Н. Лидерство бизнес-и системного аналитика на IT-рынке // Программно-техническое обеспечение автоматизированных систем. – 2021. – С. 17–22.
12. Кондуров И. В., Тушев А. Н. Технические основы системного аналитика для успешной коммуникации с командой разработки // Высокопроизводительные вычислительные системы и технологии. – 2020. – Т. 4. – №. 2. – С. 91–95.
13. Костогрызов А. И. К методам системной инженерии: вероятностные подходы к анализу процесса управления качеством системы // Современные информационные технологии и ИТ-образование. – 2022. №2. – URL: <https://cyberleninka.ru/article/n/k-metodam-sistemnoy-inzhenerii-veroyatnostnye-podhody-k-analizu-protsessa-upravleniya-kachestvom-sistemy>.
14. Анализ информационной безопасности предприятия на основе сбора данных пользователей с открытых ресурсов и мониторинга информационных ресурсов с использованием машинного обучения // T-Comm. – 2018. №10. – URL: <https://cyberleninka.ru/article/n/analiz-informatsionnoy-bezopasnosti-predpriyatiya-na-osnove-sbora-dannyh-polzovateley-s-otkrytyh-resursov-i-monitoringa>
15. Ловчиков, А. Е. Бизнес-аналитик: особенности и перспективы профессии // Огарёв-Online. – 2022. №8 (177). – URL: <https://cyberleninka.ru/article/n/biznes-analitik-osobennosti-i-perspektivy-professii>
16. Львович И. Я. Проблемы методологии проектирования интеллектуальных информационных систем // Информационные технологии в управлении, автоматизации и мехатронике. – 2020. – С. 120–123.
17. Люкевич И. Н., Горбатенко И. И., Пынзарь Е. Г. Цифровые технологии финансовых рынков: платформы технического анализа // Фундаментальные и прикладные исследования в области управления, экономики и торговли. – 2021. – С. 101–112.
18. Маркеева А. В., Гавриленко О. В. Большие данные как исследовательская технология: возможности и ограничения применения в современной управленческой практике // Общество: социология, психология, педагогика. – 2021. №12 (92). – URL:

<https://cyberleninka.ru/article/n/bolshie-dannye-kak-issledovatel'skaya-tehnologiya-vozmozhnosti-i-ogranicheniya-primeneniya-v-sovremennoy-upravlencheskoy-praktike>.

19. Михелёв В. В. Системно-объектный подход к системному анализу: особенности и преимущества // Экономика. Информатика. – 2022. №1. – URL: <https://cyberleninka.ru/article/n/sistemno-obektnyy-podhod-k-sistemnomu-analizu-osobennosti-i-preimuschestva>.

20. Ольховская И. В., Ишанходжаев М. А. Использование бизнес-интеллекта и бизнес-аналитики в организациях // Экономика и бизнес: теория и практика. – 2018. №4. – URL: <https://cyberleninka.ru/article/n/ispolzovanie-biznes-intellekta-i-biznes-analitiki-v-organizatsiyah>.

21. Панкратова Н. Д., Панкратов В. А. Роль и место системного анализа в практической деятельности // Системный анализ в проектировании и управлении. – 2019. – Т. 23. – №. 1. – С. 31–40.

22. Родионова П. Д. Применение цифровых технологий на рынке ценных бумаг // Кластеризация цифровой экономики: Глобальные вызовы. – 2020. – С. 205–209.

23. Садовский Г. Л. Применение больших данных и систем аналитики для эффективного управления проектами // Управление научно-техническими проектами. – 2020. – С. 225–228.

24. Саханевич Д. Ю., Кремин А. Е. Систематизация методов машинного обучения в целях внедрения искусственного интеллекта в социально-экономические процессы региона // Вестник Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых. Серия: Экономические науки. – 2019. – № 4 (22). – С. 57–65.

25. Субботин А. В., Тагирова, 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.

26. Системный анализ задачи моделирования естественного языка // Труды Кольского научного центра РАН. – 2021. №5 (12). – URL: <https://cyberleninka.ru/article/n/sistemnyy-analiz-zadachi-modelirovaniya-estestvennogo-yazyka>

27. Чернышева Ю. Г. Новая концепция аналитики в организации - бизнес-анализ // Учет и статистика. – 2019. №2 (54). – URL: <https://cyberleninka.ru/article/n/novaya-kontseptsiya-analitiki-v-organizatsii-biznes-analiz>.

28. Чижик В. П. Сравнительная характеристика методов фундаментального и технического анализа финансовых активов // Сибирский торгово-экономический журнал. – 2013. – №. 1 (17). – С. 49.

Щурина С. В., Данилов А. С. Искусственный интеллект как технологическая инновация развития экономики // Экономика. Налоги. Право. – 2019. – № 12 (3). – С. 125–133.

Информационное сопровождение

Электронные образовательные ресурсы

Электронные информационные ресурсы

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

1. <https://xn--80aapampemcchfmo7a3c9ehj.xn--p1ai/projects/tsifrovaya-ekonomika> – сайт

<p>2. http://www.intuit.ru/studies/professional_retraining/953/courses/214/lecture/5508/</p> <p>3. https://fpmi-edu.ru/free-de</p> <p>4. https://ru.coursera.org/specializations/gcp-data-machine-learning#courses</p> <p>5. https://www.coursera.org/specializations/big-data</p> <p>6. https://stepik.org/course/101687/promo</p> <p>7. https://mlcourse.ai/book/index.html</p>	<p>Национальной программы «Цифровая экономика»</p> <p>2. https://xn--80aapampemcchfmo7a3c9ehj.xn--p1ai/projects/tsifrovaya-ekonomika/p-iskusstvennyu-intellekt-p – сайт федерального проекта «Искусственный интеллект» Национального проекта «Цифровая экономика»</p> <p>3. https://ai.2035.university/ – сайт УНТИ 2035 «Обучение в области искусственного интеллекта»</p> <p>4. https://sprint.lt.ru/ – сайт образовательной платформы ООО «1Т»</p> <p>5. https://trends.rbc.ru/trends/tag/ai – раздел «Искусственный интеллект» на сайте РБК.</p> <p>6. https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82_(%D0%98%D0%98,_Artificial_intelligence,_AI) – раздел «Искусственный интеллект» на сайте TAdviser. Государство. Бизнес. Технологии.</p>
---	---

7.4 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
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 100Мбит; веб-браузеры, 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.