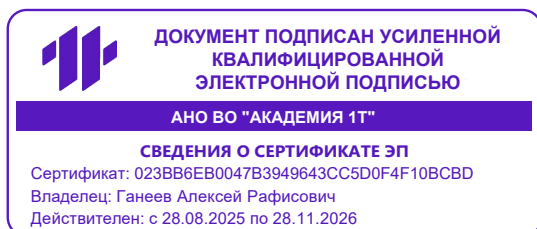


Автономная некоммерческая организация высшего образования «Академия 1Т»



«УТВЕРЖДАЮ»

Ректор АНО ВО «Академия 1Т»

/И.Р. Ганеев/



ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ ПО НАПРАВЛЕНИЮ
DATA SCIENCE

Срок обучения: 260 академических часов.

Форма обучения: очно-заочная форма обучения, осуществляемая с применением электронного обучения и дистанционных образовательных технологий без отрыва от производства.

Категория слушателей: лица, имеющие среднее профессиональное образование, высшее образование – бакалавриат (магистратура).

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ.....	3
1. ОБЩИЕ ПОЛОЖЕНИЯ	4
1.1 Цель и задачи дополнительной профессиональной программы профессиональной переподготовки	5
1.2. Характеристика вида профессиональной деятельности, трудовых функций и(или) уровней квалификации.....	8
1.3 Требования к результатам освоения дополнительной профессиональной программы профессиональной переподготовки Data Science	9
1.4 Требования к итоговой аттестации по дополнительной профессиональной программе профессиональной переподготовки Data Science	11
2. УЧЕБНЫЙ ПЛАН.....	12
3. РАБОЧИЕ ПРОГРАММЫ	17
3.1. Рабочая программа 1.	17
3.2. Рабочая программа 2.	24
4. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ.....	30
5. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК	33
6. ФОРМЫ АТТЕСТАЦИИ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ	34
6.1. Текущий контроль	34
6.2. Промежуточная аттестация	35
6.3. Итоговая аттестация	38
7. СПИСОК ЛИТЕРАТУРЫ	54

1. ОБЩИЕ ПОЛОЖЕНИЯ

Настоящая дополнительная профессиональная программа профессиональной переподготовки по направлению Data Science разработана в соответствии с требованиями приказа Министерства образования и науки Российской Федерации от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», Постановления Правительства РФ от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ». а также ФГОС ВО 38.03.05 Бизнес-информатика, ФГОС 09.03.04. Программная инженерия, квалификационных характеристик должностей (раздел «Квалификационные характеристики должностей работников, занятых на предприятиях, учреждениях, организациях»): «Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты РФ от 6 июля 2020 года №405н; «Программист», утвержденный приказом Министерства труда и социальной защиты РФ от 20 июля 2022 года N 424н.

Структура дополнительной профессиональной программы профессиональной переподготовки по направлению Data Science включает:

цель и задачи дополнительной профессиональной программы профессиональной переподготовки;

планируемые результаты обучения;

учебный план;

календарный учебный график;

рабочая программа учебных модулей;

организационно-педагогические условия реализации дополнительной профессиональной программы профессиональной переподготовки;

формы аттестации и оценочные материалы;

список литературы.

Срок освоения дополнительной профессиональной программы профессиональной переподготовки – 260 академических часов.

Срок освоения дополнительной профессиональной программы профессиональной переподготовки обеспечивает возможность достижения планируемых результатов и получение новых компетенций в соответствии с целью и задачами программы.

Форма обучения: очно-заочная форма обучения, осуществляемая с применением электронного обучения и дистанционных образовательных технологий (онлайн-вебинары и т. п.) без отрыва от производства.

1.1 Цель и задачи дополнительной профессиональной программы профессиональной переподготовки

Цель дополнительной профессиональной программы профессиональной переподготовки

Цель дополнительной профессиональной программы профессиональной переподготовки по направлению Data Science — получение слушателями компетенций, необходимых для профессиональной деятельности в области искусственного интеллекта (далее — ИИ) для разработки и применения технологических решений в области искусственного интеллекта и в смежных областях.

Задачи дополнительной профессиональной программы профессиональной переподготовки

Задачи дополнительной профессиональной программы профессиональной переподготовки по направлению Data Science:

формирование теоретических знаний, практических умений и навыков по основным принципам работы искусственного интеллекта и его видам, изучение основного инструментария для работы с данными; практическое применение методов машинного обучения для работы с большими объемами данных;

приобретение слушателями знаний и практического опыта по созданию и поддержке системы искусственного интеллекта на основе нейросетевых моделей и методов, овладение языками программирования (Python) и другими инструментами для реализации методов анализа и структурирования больших данных;

получение новых компетенций, необходимых для профессиональной деятельности в рамках реализации ФГОС ВО по направлению подготовки 38.03.05 Бизнес-информатика (приказ Минобрнауки России от 29.07.2020), 09.03.04. Программная инженерия (Приказ Минобрнауки России от 19.09.2017 N 932 (ред. от 08.02.2021)).

Содержание дополнительной профессиональной программы профессиональной переподготовки по направлению Data Science и отдельных ее компонентов (дисциплин

(модулей)) направлено на достижение цели и задач программы, планируемых результатов ее освоения.

Содержание дополнительной профессиональной программы профессиональной переподготовки Data Science построено по модульному принципу, единицами которого являются модули. Кодировка структурных единиц программы осуществляется следующим образом: в начале нумерации идёт код модуля (например, 1), далее – код раздела (например, 1.1), далее – код темы (лекции) (например, 1.1.1). Данная кодировка используется для удобства использования содержания программы в процессе обучения.

В процессе освоения дополнительной профессиональной программы профессиональной переподготовки слушатели овладеют знаниями:

- о функциях, задачах, навыках, содержании работы Data Scientist в области искусственного интеллекта;

- об истории развития и главных трендах искусственного интеллекта;

- об основных типах, источниках больших данных;

- о методах и программном инструменте технологий больших данных;

- о синтаксисе и семантике основных языков искусственного интеллекта (Python) и основных приемах программирования на них;

- об основных алгоритмах, используемых в языках программирования;

- о методах, инструментальных средствах и стандартах статистического и описательного анализа данных;

- о системах обработки и анализа больших массивов данных (SQL, NoSQL): принципы работы, функционал, язык запросов, преимущества и недостатки;

- об основных платформах данных (облачные и внутрикорпоративные), применяемых при разработке решений на основе ИИ;

- об архитектуре, принципах, процессе, стадии и методологии разработки решений на основе ИИ;

- об основных видах представления данных: табличные, временные ряды;

- о принципах, содержании и области применения массово-параллельной обработки и анализа данных для ускорения машинного обучения;

- об основах методов оптимизации;

- о технологии, методах и инструментах машинного обучения для решения профессиональных задач;

- о технологии распределенной обработки данных;

- о принципах и содержании работы с распределенными кластерными системами;

об основных уровнях представления данных;

об архитектуре современных нейронных сетей и их использовании для решения профессиональных задач.

Учебный план дополнительной профессиональной программы профессиональной переподготовки определяет перечень, трудоемкость, последовательность и распределение изучаемых тем дисциплин (модулей), устанавливает формы организации учебного процесса и их соотношение (лекции, практические занятия, самостоятельная работа), определяет формы аттестации.

Дополнительная профессиональная программа профессиональной переподготовки Data Science включает планируемые результаты обучения и требования к итоговой аттестации обучающихся.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Планируемые результаты обучения по дополнительной профессиональной программе профессиональной переподготовки Data Science направлены на развитие профессиональных компетенций слушателей, необходимых для выполнения нового вида профессиональной деятельности. В планируемых результатах отражается преемственность с профессиональными стандартами и квалификационными характеристиками должностей работников ИТ-сферы.

1.2. Характеристика вида профессиональной деятельности, трудовых функций и(или) уровней квалификации

Наименование базовой группы, должности (профессии) или специальности, согласно Единому квалификационному справочнику должностей (раздел «Квалификационные характеристики должностей работников, занятых на предприятиях, учреждениях, организациях»): «Начальник вычислительного (информационно-вычислительного) центра», «Начальник отдела информации».

Вид профессиональной деятельности:

Создание и применение технологий больших данных.

Требования к образованию и обучению слушателей, опыту практической работы, трудовая функция и трудовые действия.

Требования к образованию и обучению слушателей: среднее профессиональное образование, высшее образование – бакалавриат (магистратура).

Требования к опыту практической работы слушателей: без опыта.

Трудовая функция:

планирование и организация аналитических работ с использованием технологий больших данных;

подготовка данных для проведения аналитических работ по исследованию больших данных;

анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры;

управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации;

управление разработкой продуктов, услуг и решений на основе больших данных;

разработка и внедрение новых методов и технологий исследования больших данных.

Трудовые действия:

разработка, обсуждение и утверждение содержания аналитических работ с использованием технологий больших данных;

определение источников больших данных для анализа, идентификация внешних и внутренних источников данных для проведения аналитических работ;

разработка, обсуждение и утверждение содержания аналитических работ с использованием технологий больших данных;

получение и фильтрация больших объемов данных из гетерогенных источников;

оценка соответствия набора данных предметной области и задачам аналитических работ;

выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ;

разработка, проверка, оценка используемых моделей больших данных;

адаптация и развертывание моделей больших данных в предметной среде;

подготовка отчета по результатам аналитических работ с использованием технологий больших данных;

консультирование заказчика по результатам аналитических работ с применением технологий больших данных;

мониторинг эффективности работы аналитики больших данных;

формирование предложений по использованию результатов анализа больших данных: рассылка, создание приложений, оптимизация процессов;

формирование предложений по развитию существующей методологической и технологической инфраструктуры анализа больших данных на основе выполненных работ.

1.3 Требования к результатам освоения дополнительной профессиональной программы профессиональной переподготовки Data Science

Компетенции, подлежащие совершенствованию в результате освоения дополнительной профессиональной программы профессиональной переподготовки Data Science

Общие компетенции (далее – ОК):

способность организовывать собственную деятельность, определять методы решения профессиональных задач, оценивать их эффективность и качество (ОК-2);

способность осуществлять поиск, анализ и оценку информации, необходимой для постановки и решения профессиональных задач, профессионального и личностного развития (ОК-4);

способность строить профессиональную деятельность с соблюдением регулирующих ее правовых норм (ОК-5).

Общепрофессиональные компетенции (далее – ОПК):

готовность к профессиональной деятельности в соответствии с нормативно-правовыми актами сферы Data Science (ОПК-1);

способность использовать алгоритмы и программы, современные информационные технологии, методы и средства контроля, диагностики и управления, пригодные для практического применения в сфере своей профессиональной деятельности (ОПК-2);

способность оценивать возможные риски и принимать решения в нестандартных ситуациях (ОПК-3).

Профессиональные компетенции (далее – ПК):

способность классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта (ПК-1);

способность применять методы и программные средства автоматизированного логического вывода и автоматизированной проверки гипотез (ПК-2);

способность участвовать в процессе создания систем искусственного интеллекта на различных этапах жизненного цикла в качестве эксперта и ключевого пользователя (ПК-3);

способность разрабатывать и применять методы машинного обучения для решения задач (ПК-4);

способность использовать инструментальные средства для решения задач машинного обучения (ПК-5);

способность осуществлять сбор и подготовку данных для систем искусственного интеллекта (ПК-6);

способность создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта (ПК-7).

Новые компетенции, формирующиеся в результате освоения дополнительной профессиональной программы профессиональной переподготовки Data Science» (далее – НК):

умение создавать архитектуру в области ИИ для разработки и применения технологических решений (НК-1);

умение использовать системы искусственного интеллекта на основе нейросетевых моделей и методов (НК-2);

умение создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов (НК-3);

умение использовать цифровые платформы анализа данных для решения профессиональных задач (НК-4).

1.4 Требования к итоговой аттестации по дополнительной профессиональной программе профессиональной переподготовки Data Science

Освоение дополнительной профессиональной программы профессиональной переподготовки Data Science завершается итоговой аттестацией. Итоговая аттестация проходит в форме решения практико-ориентированных задач (кейсов). Итоговая аттестация выявляет теоретическую и практическую подготовку обучающегося в соответствии с целью и задачами программы.

Слушатели допускаются к итоговой аттестации только после изучения программы в определенном учебным планом объеме.

Слушатели, успешно освоившие дополнительную профессиональную программу профессиональной переподготовки Data Science и прошедшие итоговую аттестацию, получают документ о квалификации установленного образца – диплом о профессиональной переподготовке.

2. УЧЕБНЫЙ ПЛАН

Учебный план определяет перечень изучаемых тем с обозначением их трудоемкости, объема, последовательности и сроков изучения, устанавливает формы организации учебного процесса и их соотношение (лекции, практические занятия, самостоятельная работа), а также форму аттестации.

Учебный план дополнительной профессиональной программы профессиональной переподготовки по направлению Data Science представлен в таблице.

Форма обучения: очно-заочная с применением электронного обучения и дистанционных образовательных технологий без отрыва от производства.

Общая трудоемкость: 260 академических часов.

Название дисциплины (модуля)	Общая трудоемкость	Аудиторные занятия, акад. час		Самостоятельная работа, акад. час.	Текущий контроль	Форма аттестации
		Лекции	Практика			
Модуль 1. Базовый	142	34	76	32		Решение практико-ориентированных задач (кейсов)
Раздел 1.1. Введение	58	18	28	12		
1.1.1 Введение в профессию Data Science.	8	2	4	2	Тест	
1.1.2 Введение в Python	10	4	4	2	Тест	
1.1.3 Введение в основные библиотеки Python для работы с данными	8	2	4	2	Тест	
1.1.4 Статистический и описательный анализ данных	12	4	6	2	Тест	
1.1.5 Введение в основные библиотеки Python для визуализации данных. Matplotlib, Seaborn, Plotly	10	4	4	2	Тест	
1.1.6 EDA — разведочный анализ данных. Как найти особенности в данных	10	2	6	2	Тест	
Раздел 1.2. Погружение в ML	80	16	48	16		

1.2.1 Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей	10	2	6	2	Тест
1.2.2 Линейная алгебра, работа с numpy, принцип работы классических моделей ML	10	2	6	2	Тест
1.2.3 Погружение в ML. Обучение с учителем	10	2	6	2	Тест
1.2.4 Способы улучшить качество модели. Feature engineering	10	2	6	2	Тест
1.2.5 Погружение в ML. Обучение без учителя	10	2	6	2	Тест
1.2.6 Погружение в ML. Временные ряды	10	2	6	2	Тест
1.2.7 Погружение в ML. Рекомендательные системы	10	2	6	2	Тест
1.2.8. Развертывание модели в производство. Pipelines. (ML в продакшене)	10	2	6	2	Тест
Промежуточная аттестация по модулю 1	4			4	
Модуль 2. Профильный	110	20	56	34	

Раздел 2.1. Нейронные сети	68	12	36	20		Решение практико-ориентированных задач (кейсов)
2.1.1 Введение в нейронные сети	12	2	6	4	Тест	
2.1.2 Сверточные и полносвязные нейронные сети	12	2	6	4	Тест	
2.1.3 CV - компьютерное зрение	12	2	6	4	Тест	
2.1.4 NLP – обработка естественного языка	12	2	6	4	Тест	
2.1.5 GAN - Генеративные модели	10	2	6	2	Тест	
2.1.6 RL - обучение с подкреплением	10	2	6	2	Тест	
Раздел 2.2. Основы Big Data	38	8	20	10		
2.2.1. Введение в Базы данных и теорию хранения данных	8	2	4	2	Тест	
2.2.2. Реляционные базы данных	8	2	4	2	Тест	
2.2.3. Введение в SQL	12	2	6	4	Тест	
2.2.4. Введение BI-системы	10	2	6	2	Тест	
Промежуточная аттестация по модулю 2	4			4		

Итоговая аттестация	8			8		Решение практико-ориентированных задач (кейсов)
Итого	260	54	132	74		

3. РАБОЧИЕ ПРОГРАММЫ

3.1. Рабочая программа 1.

Модуль 1. Базовый

1.1 Введение.

1.1.1 Введение в профессию Data Science (лекции – 2 ч., практические занятия – 4 ч., самостоятельная работа – 2 ч.)

Кто такие специалисты Data Science, какие задачи решают. Какие технологии используются в DS. Основные термины и понятия машинного обучения.

В результате обучающийся должен

уметь:

ориентироваться в используемых технологиях, основных терминах и понятиях машинного обучения;

знать:

задачи, которые решают специалисты Data Science;

используемые технологии Data Science;

основные термины и понятия машинного обучения.

1.1.2. Введение в Python (лекции – 4 ч., практические занятия – 4 ч., самостоятельная работа – 2 ч.)

Основы языка программирования Python и его особенности в работе с данными. Структуры данных Python. Работы со списками, срезами и циклами. Популярные площадки для работы с данными — Google Colab и Kaggle.

В результате обучающийся должен

уметь:

анализировать структуры данных Python;

работать со списками, срезами и циклами;

устанавливать интерпретаторы и окружения для работы с Python — VScode, Pycharm, Jupyter;

работать с Google Colab и Kaggle;

знать:

особенности языка Python;
структуру данных Python;
алгоритмы работы со списками и циклами.

1.1.3 Введение в основные библиотеки Python для работы с данными (лекции – 2 ч., практические занятия – 4 ч., самостоятельная работа – 2 ч.)

Основные библиотеки Python для работы с данными, включая Pandas и NumPy. Типы данных библиотек и как с ними работать. Возможности агрегирующих функций. Соединение таблиц, обработка пропусков и фильтрация данных.

В результате обучающийся должен

уметь:

работать с табличными данными;
работать с библиотеками;
создавать собственные функции lambda для Pandas;
обрабатывать пропуски и фильтровать данные;

знать:

основные библиотеки Python для работы с данными, включая Pandas и NumPy;
типы данных библиотек;
возможности агрегирующих функций;
правила соединения таблиц, обработки пропусков и фильтрации данных.

1.1.4 Статистический и описательный анализ данных (лекции – 4 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Методы анализа и проверки гипотез на основе статистических данных с помощью языка программирования Python. Методы обработки данных, основные понятия и принципы статистики, а также методы проверки статистических гипотез. Различные методы анализа данных и как их применять при проверке гипотез.

В результате обучающийся должен

уметь:

анализировать и проверять гипотезы на основе статистических данных с помощью языка программирования Python;

применять методы обработки данных, методы проверки статистических гипотез;

применять методы анализа данных при проверке гипотез;

знать:

методы анализа и проверки гипотез на основе статистических данных с помощью языка программирования Python;

методы обработки данных, основные понятия и принципы статистики;

методы проверки статистических гипотез;

методы анализа данных и правила их применения при проверке гипотез.

1.1.5 Введение в основные библиотеки Python для визуализации данных. Matplotlib, Seaborn, Plotly (лекции – 4 ч., практические занятия – 4 ч., самостоятельная работа – 2 ч.)

Визуализация данных в Python. Основные библиотеки Matplotlib, Seaborn, позволяющие познакомиться с базовыми инструментами визуализации, настраиваемыми графиками и различными типами диаграмм. Интерактивная визуализация данных и создание интерактивных графических отчетов в библиотеке Plotly.

В результате обучающийся должен

уметь:

визуализировать данные в Matplotlib, Seaborn;

настраивать графики и различные типы диаграмм;

создавать интерактивных графических отчетов в библиотеке Plotly;

знать:

библиотеки Matplotlib, Seaborn, Plotly;

возможности интерактивной визуализации данных и создание интерактивных графических отчетов;

базовые инструменты визуализации, настраиваемые графиками и различными типами диаграмм.

1.1.6 EDA — разведочный анализ данных. Как найти особенности в данных. (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Использование языка программирования Python и его библиотек для анализа данных и презентации полученных результатов. Исследовательский анализ данных (EDA). Как анализ данных может повлиять на принятие решений.

В результате обучающийся должен уметь:

использовать язык программирования Python и его библиотеки для анализа данных и презентации полученных результатов;

проводить исследовательский анализ данных (EDA);

знать:

правила использования язык программирования Python и его библиотеки для анализа данных и презентации полученных результатов;

алгоритм исследовательского анализа данных (EDA);

как анализ данных может повлиять на принятие решений.

1.2. Погружение в ML.

1.2.1. Введение в ML. Основные задачи, виды классических моделей, метрики качества моделей (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Виды классических моделей (линейные, деревья). Метрики качества моделей. Ансамбли алгоритмов, градиентный спуск, бустинг. Разработка моделей ML – библиотека scikit-learn.

В результате обучающийся должен уметь:

составлять метрики качества моделей;

разрабатывать модель ML – библиотеку scikit-learn

знать:

понятия основных классических моделей;

метрики качества моделей;

ансамбли алгоритмов, градиентный спуск, бустинг.

1.2.2. Линейная алгебра, работа с NumPy, принцип работы классических моделей ML (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Линейная алгебра, работа с массивами NumPy, математика ML. Работа с матрицами и линейная регрессия изнутри. Производная, градиент и оптимизация. Разработка модели ML – на основе функции, нахождение производной и градиентный спуск. Градиентная оптимизация.

В результате обучающийся должен

уметь:

работать с матрицами;

осуществлять разработку модели ML – на основе функции;

находить производную и градиентный спуск;

проводить градиентную оптимизацию;

знать:

основы линейной алгебры и математики ML;

алгоритмы работы с массивами NumPy;

алгоритмы работы с матрицами и линейная регрессия изнутри;

как находить производную, градиент и оптимизацию;

как разрабатывать модель ML;

понятие градиентной оптимизации.

1.2.3. Погружение в ML. Обучение с учителем (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Основные популярные модели ML – ансамблевые модели, градиентный бустинг, линейные алгоритмы. Основные проблемы, связанные с обучением моделей и получение лучшего качества моделей. Переобучение и недообучение моделей.

В результате обучающийся должен

уметь:

осуществлять обучение моделей и получение лучшего качества моделей;

отличать переобучение и недообучение моделей;

знать:

основные популярные модели ML;

основные проблемы, связанные с обучением моделей и получение лучшего качества моделей:

механизмы переобучения и недообучения моделей.

1.2.4. Способы улучшить качество модели. Feature engineering. (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Основные проблемы, связанные с обучением моделей и получение лучшего качества моделей. Переобучение и недообучение моделей. Подготовка/кодирование признаков, разделение выборок train/test и борьба с дисбалансом классов. Кросс-валидация.

В результате обучающийся должен

уметь:

осуществлять обучение моделей и получение лучшего качества моделей;

подготавливать/кодировать признаки, разделять выборки train/test и бороться с дисбалансом классов;

проводить кросс-валидацию;

знать:

основные проблемы, связанные с обучением моделей и получение лучшего качества моделей;

переобучение и недообучение моделей;

механизмы подготовки/кодирования признаков, разделения выборок train/test и борьбы с дисбалансом классов;

кросс-валидацию.

1.2.5. Погружение в ML. Обучение без учителя (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Обучение без учителя – уменьшение размерности, поиск закономерностей/аномалий, кластеризация.

В результате обучающийся должен

уметь:

уменьшать размерность;

искать и находить закономерности/аномалии;

проводить кластеризацию;

знать:

как уменьшить размерность;

алгоритм поиска закономерностей/аномалий;

кластеризацию.

1.2.6. Погружение в ML. Временные ряды (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Новые типы задач в целях ML – прогнозирование временных рядов — основные методы, способы валидации, метрики.

В результате обучающийся должен

уметь:

прогнозировать временные ряды;

знать:

прогнозирование временных рядов — основные методы, способы валидации, метрики.

1.2.7. Погружение в ML. Рекомендательные системы. (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.).

Новые типы задач в целях ML – рекомендательные системы — основные методы, способы валидации, метрики.

В результате обучающийся должен

уметь:

прогнозировать временные ряды;

знать:

прогнозирование временных рядов — основные методы, способы валидации, метрики.

1.2.8. Развертывание модели в производство. Pipelines. (ML в продакшене) (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

Основные проблемы и востребованность данных задач со стороны бизнеса. Все шаги создания сервиса и интеграции ML-модели в эксплуатацию.

В результате обучающийся должен

уметь:

создавать сервис и осуществлять интеграцию ML-модели в эксплуатацию;

знать:

основные проблемы и востребованность данных задач со стороны бизнеса;

все шаги создания сервиса и интеграции ML-модели в эксплуатацию.

3.2. Рабочая программа 2.

Модуль 2. Профильный.

2.1. Нейронные сети.

2.1.1. Введение в нейронные сети (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 4 ч.)

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

В результате обучающийся должен

уметь:

применять алгоритм обратного распространения ошибки для обучения нейронных сетей;

осуществлять производную и оптимизацию нейронных сетей;

знать:

понятие нейронной сети;

алгоритм обратного распространения ошибки для обучения нейронных сетей;

основные архитектуры и задачи нейронных сетей.

2.1.2. Сверточные и полносвязные нейронные сети (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 4 ч.)

Продвинутое обучение нейронных сетей. Архитектура CNN, свёрточные нейронные сети. Разница между моделями и как устроена архитектура. Работа с GPU.

В результате обучающийся должен

уметь:

осуществлять продвинутое обучение нейронных сетей;
реализовывать архитектуру CNN, свёрточные нейронные сети;
работать с GPU;

знать:

основы продвинутого обучения нейронных сетей;
архитектуру CNN, свёрточные нейронные сети;
разница между моделями и как устроена архитектура;
алгоритмы работы с GPU.

2.1.3. CV — компьютерное зрение (лекции — 2 ч., практические занятия — 6 ч., самостоятельная работа — 4 ч.)

Классификация, сегментация изображений. Распознавание объектов на изображении. Использование CV в видео. Распознавание текста на изображении. Способы оптимизации.

В результате обучающийся должен

уметь:

проводить классификацию, сегментацию изображений;
распознавать объекты на изображении и текст на изображении;
применять алгоритмы использования CV в видео;
применять способы оптимизации;

знать:

классификацию, сегментацию изображений;
как распознать объекты на изображении;
алгоритмы использования CV в видео;
как распознать текст на изображении;
способы оптимизации.

2.1.4. NLP — обработка естественного языка (лекции — 2 ч., практические занятия — 6 ч., самостоятельная работа — 4 ч.)

Основные задачи и для чего используется NLP. Что такое векторное представление слов, мешок слов и N-граммы, TF-IDF, эмбендинги. Виды нейронных сетей для задач NLP. Использование BERT для построения эмбендинга слов.

В результате обучающийся должен

уметь:

использовать векторное представление слов, мешок слов и N-граммы, TF-IDF, эмбендинги;

различать виды нейронных сетей для задач NLP;

использовать BERT для построения эмбендинга слов;

знать:

основные задачи и для чего используется NLP;

что такое векторное представление слов, мешок слов и N-граммы, TF-IDF, эмбендинги;

виды нейронных сетей для задач NLP;

алгоритмы использования BERT для построения эмбендинга слов.

2.1.5. GAN — Генеративные модели (лекции — 2 ч., практические занятия — 6 ч., самостоятельная работа — 2 ч.)

Введение в генерацию текстов. Текстовая генерация с помощью трансформеров.

В результате обучающийся должен

уметь:

осуществлять генерацию текстов;

осуществлять текстовую генерацию с помощью трансформеров;

знать:

понятие генеративной модели;

механизм генерации текстов;

алгоритм текстовой генерации с помощью трансформеров.

2.1.6. RL — обучение с подкреплением (лекции — 2 ч., практические занятия — 6 ч., самостоятельная работа — 2 ч.)

Reinforcement Learning – что такое обучение с подкреплением, кто такой агент. Создание собственной модели RL.

В результате обучающийся должен

уметь:

создавать собственную модель RL;

знать:

что такое обучение с подкреплением;

кто такой агент в Reinforcement Learning;

механизм создания собственной модели RL.

2.2. Основы Big Data

2.2.1 Введение в Базы данных и теорию хранения данных. (лекции – 2 ч., практические занятия – 4 ч., самостоятельная работа – 2 ч.)

Принципы организации баз данных, их структура и способы хранения информации. Основные термины: таблица, поле, запись, ключ и другие. Преимущество использования баз данных перед простыми файлами и таблицами в электронных таблицах.

В результате обучающийся должен

уметь:

создавать базы данных по заданным параметрам

знать:

понятие базы данных, основные термины и принципы организации.

2.2.2. Реляционные базы данных (лекции – 2 ч., практические занятия – 4 ч., самостоятельная работа – 2 ч.)

Основы реляционных баз данных. Структура и функциональность реляционных баз данных, а также ключевые концепции, такие как первичный и внешний ключ. Как реляционные базы данных могут помочь в управлении, хранении и анализе данных. Настройка и оптимизация базы данных для различных задач и бизнес-целей.

В результате обучающийся должен

уметь:

настраивать реляционную базу данных для различных целей;

оптимизировать базу данных;

знать:

понятие реляционной базы данных;

ключевые концепции баз данных;

основные команды базы данных.

2.2.3. Введение в SQL (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 4 ч.)

SQL как инструмент, позволяющий извлекать, добавлять, изменять и удалять данные, а также управлять структурой базы данных. Основы SQL, включая работу с данными, использование условий и группировку, а также применение функций агрегирования и соединения таблиц.

В результате обучающийся должен

уметь:

создавать запросы и подзапросы к реляционной СУБД Postgres;

осуществлять операции соединения в реляционной СУБД Postgres;

осуществлять оптимизацию SQL запросов;

знать:

основные команды SQL, диалекты;

функции агрегирования и соединения таблиц.

2.2.4. Введение в BI-системы (лекции – 2 ч., практические занятия – 6 ч., самостоятельная работа – 2 ч.)

BI-системы как инструменты для анализа данных и принятия управленческих решений в компаниях. Функции BI-систем: сбор, обработка и анализ данных из различных источников. Источники: базы данных, системы ERP, системы CRM и другие приложения компаний. Основные компоненты BI-систем: отчеты и инструменты управления.

В результате обучающийся должен

уметь:

применять инструменты BI-системы;

работать с различными источниками данных (базы данных, системы ERP, системы CRM);

знать:

основные понятия, функции BI-систем, их целевое назначение;

источники данных;

основные компоненты BI-систем: отчеты и инструменты управления.

4. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ

Специфика организационных действий и педагогических условий

Для достижения планируемых результатов обучение строится с использованием следующих:

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

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

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

Кадровое обеспечение программы (преподавательский состав)

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

Требования к представителям образовательной организации высшего образования и научной организации: высшее образование, ученая степень, стаж научно-педагогической работы более трех лет.

Требования к представителям компаний со стажем работы в области искусственного интеллекта и в смежных областях: опыт решения практических задач с использованием технологий искусственного интеллекта более 3 лет в профильной компании или в профильном подразделении.

Материально-технические условия реализации программы

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

	интерфейс Fast Ethernet 100 Мбит; веб-браузеры Google Chrome, Mozilla Firefox, Opera, Microsoft Edge, Яндекс.Браузер и др. Anaconda 2.7 или 3.5 Доступ к облачным вычислительным ресурсам.
--	--

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

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

Общие требования к организации учебного процесса

Общие требования к организации учебного процесса определяются локальными нормативными актами образовательной организации.

В начале обучения слушатели знакомятся с порядком организации работы с модулями программы, учебным планом и формами аттестации.

В процессе обучения слушатели обеспечиваются текстовыми, видео и презентационными материалами по всем модулям программы.

Образовательный процесс формируется на основе принципа синхронизации теории и практики, что достигается преимущественно практико-ориентированным характером обучения.

Образовательная деятельность слушателей предусматривает следующие виды учебных занятий и учебных работ: видео-лекции, вебинары, практические занятия, самостоятельная работа.

Реализация программы предполагает:

- возможность проведения практических занятий в формате вебинаров;
- предоставление поэтапного доступа слушателей к усваиваемым разделам модуля и информационным материалам;
- систему дистанционного учета результатов освоения модуля, в том числе проведения промежуточной аттестации по результатам освоения разделов модуля;
- обеспечение свободного доступа педагогов к средствам информационно-дистанционных технологий, с оборудованным рабочим местом, оснащенным персональным

компьютером, веб-камерой, динамиком, микрофоном для проведения консультаций и лекционных занятий;

– учебно-методическую документацию и материалы по тематике модуля, в том числе для лекционных занятий и самостоятельной работы; содержание материалов представлено на специализированной онлайн-платформе.

Каждый обучающийся получает возможность доступа к специализированной онлайн-платформе курсов, которая обеспечивает возможность индивидуального обучения для каждого слушателя из любой точки, имеющей доступ к сети Интернет.

Слушатели в соответствии с календарным планом-графиком обучения изучают материалы курса, выполняют практические задания и самостоятельные работы, проходят промежуточную и итоговую аттестации. Преподаватели в соответствии с календарным планом-графиком обучения проводят практические занятия, контролируют выполнение практических и самостоятельных работ, заданий промежуточной и итоговой аттестации, оперативно направляя их на доработку слушателям в случае такой необходимости. Консультирование и поддержка слушателей преподавателями осуществляется в течение всего периода обучения в онлайн и оффлайн форматах. Используемые каналы для коммуникации: общение в ходе проведения занятий, по электронной почте, групповой чат в Telegram-канале.

5. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Объем программы – 260 часов.

Продолжительность обучения – 3 месяца.

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

Режим занятий: 3-4 часа в день.

Модули и разделы программы	Сроки (учебных недель)
Модуль 1. Базовый	
Раздел 1.1. Введение	3
Раздел 1.2. Погружение в ML	4
Модуль 2. Профильный	
Раздел 2.1. Нейронные сети	3
Раздел 2.2. Основы Big Data	2

6. ФОРМЫ АТТЕСТАЦИИ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Слушатели допускаются к итоговой аттестации после изучения дополнительной профессиональной программы профессиональной переподготовки Data Science в полном объеме, предусмотренным учебным планом.

Учебным планом предусмотрены следующие формы контроля освоения теоретического и практического материала рабочей программы дополнительной профессиональной программы профессиональной переподготовки Data Science: текущий контроль, промежуточная аттестация, итоговая аттестация.

6.1. Текущий контроль

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

Примеры тестов.

Модуль 1. Базовый

Раздел 1.2 Основы хранения и обработки данных

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

Вопросы:

1. Какой вид задач решает алгоритм машинного обучения, называемый «классификация»?
 - а) Прогнозирование числового значения
 - б) Выделение групп похожих объектов
 - с) Определение принадлежности объекта к определенному классу**
 - д) Разбиение выборки на обучающую и тестовую

2. Какой из этих методов не относится к классическим моделям машинного обучения?
 - а) Линейная регрессия
 - б) Решающее дерево
 - с) Нейронные сети**
 - д) Логистическая регрессия

3. Какая метрика используется для оценки качества моделей классификации?

- a) RMSE
- b) R^2
- c) F1-score**
- b) MSE

4. Какая из этих задач является задачей регрессии?

- a) Определение наличия определенного объекта на изображении
- b) Кластеризация пользователей по просмотру фильмов на онлайн-платформе
- c) Прогнозирование цены на недвижимость**
- d) Определение тональности отзывов о товарах

5. Какой метод машинного обучения решает задачу поиска похожих объектов?

- a) Линейная регрессия
- b) Решающие деревья
- c) Метод опорных векторов
- d) К-ближайших соседей**

6.2. Промежуточная аттестация

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

Примеры практико-ориентированных задач (кейсов)

Модуль 1

Задание 1

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

1. Загрузить данные, обработать ошибки в данных – заполнить либо удалить пропуски, удалить дубликаты.
2. Провести EDA-анализ данных. Построить графики – выбрать оптимальные для числовых и категориальных признаков. Найти возможные закономерности в данных. Провести статистический анализ данных – сравнить выборки, где целевая переменная 1 и где 0.

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

4. Обучить модель логистической регрессии.

5. Обучить модель дерева решений.

6. Обучить модель решающих деревьев.

7. Сравнить качества моделей и выбрать лучшую модель для предсказания.

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

Задание 2

Спроектируйте схему данных и постройте аналитическую витрину:

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

Критерии и шкала оценивания:

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

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

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

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

Максимально возможное число баллов за работу – 10:

не менее 9 баллов – «отлично»;

7–8 баллов – «хорошо»;

5–6 баллов – «удовлетворительно»;

0–4 балла – «неудовлетворительно».

Модуль 2

Задание 1

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

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

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

Произвести анализ данных с помощью метода понижения размерности.

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

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

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

Сравнить качества моделей и выбрать лучшую модель для предсказания.

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

Цель: достичь метрики `RMSE` менее 2000.

Задание 2

Кластеризация новостей

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

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

Векторизировать текст с помощью любых библиотек Python.

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

Сравнить результат кластеризации с фактическим разбиением по темам на сайте.

Произвести аналитику новостей. Ссылку на репозиторий с кодом прислать в чат с преподавателем.

Критерии и шкала оценивания:

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

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

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

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

Максимально возможное число баллов за работу – 10:

не менее 9 баллов – «отлично»;

7–8 баллов – «хорошо»;

5–6 баллов – «удовлетворительно»;

0–4 балла – «неудовлетворительно».

6.3. Итоговая аттестация

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

Практико-ориентированные задачи (кейсы) для итоговой аттестации предоставлены организациями, работающими в области искусственного интеллекта и смежных областях (ООО «Альмира»), а также организациями, применяющими технологии интеллектуальной обработки данных для решения своих производственных задач (ООО «Консист»).

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

Описание задачи:

Даны табличные данные о свойствах красного вина.

Задача – разработать алгоритм прогнозирования качества вина на основе характеристик.

Этапы задачи:

1. Провести очистку и предобработку данных.
2. Провести анализ данных – исследовательский и статистический анализ данных, рассмотреть распределение качества вина по характеристикам.
3. Выбрать алгоритм машинного обучения – классификатор или регрессия.
4. Обучить модель и выбрать наилучшую.
5. Выделить самые важные признаки при формировании модели.

6. Проверить выбранную модель на адекватность (проверка со среднего значения предсказания).

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

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

Описание задачи:

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

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

Этапы задачи:

1. Спроектировать БД для хранения такого типа данных.
2. Провести анализ данных – рассмотреть дисбаланс классов.
3. Разработать сверточную нейронную сеть для классификации изображений цветов – добиться наилучшей метрики ассигасы.
4. Обучить модель и сохранить, проверить качество модели на новых данных.
5. Разработать сервис для автоматической классификации изображений.

В качестве результата работы программного продукта необходимо разработать контейнер docker и выложить результат на github.

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

Описание задачи:

Дан датасет с негативными и положительными комментариями.

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

Этапы задачи:

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

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

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

Описание задачи:

Нужно обучить генеративную модель для генерации текстов поздравлений.

Задача – разработать модель и выложить рабочую модель.

Этапы задачи:

1. Спроектировать БД для хранения запросов и результатов запросов.
2. Обучить генеративную модель.
3. Автоматизировать и максимально ускорить работу модели.
4. Разработать dashboard с результатами запросов модели – внутренней отладке и аналитике работы модели.
5. Разработать чат-бота, где пользователь отправляет «затравку» — о чем поздравление, а результат – полный текст поздравления.

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

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

Описание задачи:

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

Этапы задачи:

1. Спроектировать БД для хранения новостей по их источникам, а также для хранения результатов работы модели.
2. Спарсить все новости за месяц из выбранных источников.
3. Провести полный анализ текстов – кластеризировать тексты, выделить основные темы из результатов работы модели машинного обучения. Сохранить результаты в БД.
4. Автоматизировать работу сбора данных и работы модели через Airflow.
5. Сформировать витрину данных.
6. Разработать docker-контейнер с полным циклом работы модели.

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

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

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

Этапы задачи:

1. Провести EDA предоставленных данных по заемщикам.
2. Собрать полный отчет в Jupyter notebook – основные закономерности в данных и статистический анализ. Сформировать витрину данных.
3. Предобработать данные – очистить от дубликатов/пропусков или заполнить пропуски.
4. Обучить модель машинного обучения на основе данных.
5. Насытить данные для лучше работы модели.
6. Улучшить качество модели с помощью подбора гиперпараметров.
7. Сохранить модель для будущей работы.
8. Разработать контейнер с автоматической обработкой данных и обучением модели.

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

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

Цель: разработать модель прогнозирования спроса.

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

Этапы задачи:

1. **Сбор данных.** Необходимо собрать данные о продажах товара за определенный период времени (например, год). Для этого можно использовать любой публичный источник данных. Важно, чтобы данные были представлены в удобном формате и содержали достаточное количество информации о продажах товара (например, количество проданных единиц товара в каждый день/неделю/месяц и цена за единицу товара).
2. **Подготовка данных.** Данные, собранные на предыдущем этапе, нужно подготовить для дальнейшего анализа. Необходимо провести первичную обработку данных, оценить их качество и, при необходимости, произвести очистку данных от выбросов, ошибок и пропусков.
3. **Анализ данных.** После подготовки данных необходимо провести исследовательский анализ данных, чтобы понять характеристики и зависимости между

переменными. На этом этапе можно использовать статистические методы анализа данных, построить графики для лучшего понимания данных.

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

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

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

Результат работы программного продукта

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

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

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

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

Цель: разработать модель генерации картинок.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

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

4. **Развитие и тестирование модели.** После нужно использовать генератор для генерации новых изображений. Нужно поэкспериментировать с размером шума и другими

параметрами, чтобы получить нужные изображения. Готовую модель необходимо сохранить и разработать небольшой GUI (любая готовая библиотека, можно gradio).

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

Результат работы программного продукта

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

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

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

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

Этапы разработки:

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

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

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

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

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

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

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

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

Цель: разработать модель распознавания рукописного ввода OCR.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

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

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

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

Результат работы программного продукта

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

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

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

Цель: разработать рекомендательной модели для подбора фильмов.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

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

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

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

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

Результат работы программного продукта

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

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

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

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

Этапы задачи:

1. Провести EDA предоставленных данных по плательщикам.
2. Собрать полный отчет в Jupyter notebook – основные закономерности в данных и статистический анализ. Сформировать витрину данных.

3. Обучить модель для поиска аномалий временного ряда плательщика.

4. Сохранить модель с лучшими характеристиками.

5. Разработать контейнер с автоматической обработкой данных и обучением модели.

Результат работы программного продукта

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

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

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

Цель: обучение нейронной сети для распознавания лиц в организации.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. **Хранение данных.** Разработать архитектуру базы данных для хранения информации о людях и биометрии лиц.

2. **Обработка данных.** Обработать данные для задачи биометрии – векторизировать изображения в тензоры.

3. **Разработка модели.** Разработать нейронную сеть и провалидировать модель.

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

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

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

Результат работы программного продукта

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

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

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

Цель: разработать модель генерации картинок.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

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

Результат работы программного продукта

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

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

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

Цель: разработать модель генерации музыки.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

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

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

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

Результат работы программного продукта

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

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

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

Цель: обучение нейронной сети для распознавания лиц в организации.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

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

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

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

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

Результат работы программного продукта

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

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

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

Цель: разработать рекомендательной модели для прогнозирования погоды.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

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

3. Разработка модели. Разработать модель прогнозирования временного ряда.

4. Оценка и тестирование. Протестировать работу модели.

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

Результат работы программного продукта

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

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

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

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

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

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

Результат работы программного продукта

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

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

Цель: обучение модели для поиска оптимального маршрута.

Этапы задачи:

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

8. Разработать простой веб-интерфейс для быстрого доступа к модели.

Результат работы программного продукта

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

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

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

Цель: обучение модели для оптимизации процессов бизнеса.

Этапы задачи:

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

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

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

Цель: обучение нейронной сети для распознавания номеров машин.

Для разработки модели необходимо выполнить следующие задачи.

Этапы задачи:

1. Обработка данных. Обработать данные для задачи сегментации и определения значений на номере.
2. Разработка модели. Разработать нейронную сеть (OpenCV или любую другую модель).

3. Разработка БД. Разработать архитектуру БД для хранения номеров и машин, а также изображений.

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

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

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

Результат работы программного продукта

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

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

Критерии оценивания, шкала оценивания

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

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

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

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

Максимально возможное число баллов – 10:

не менее 9 баллов – «отлично»;

7–8 баллов – «хорошо»;

5–6 баллов – «удовлетворительно»;

0–4 балла – «неудовлетворительно».

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

Знать:

- функции, задачи, навыки, содержание работы специалиста Data Science;
- определения, историю развития и главные тренды искусственного интеллекта;
- основные типы, источники больших данных, проблемы безопасности;
- методы и программный инструментарий технологий больших данных;
- синтаксис и семантику основных языков искусственного интеллекта (Python) и основные приемы программирования на них;
- основные алгоритмы, используемые в языках программирования;

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

Уметь:

- применять системы обработки и анализа больших массивов данных;
- применять системы контроля версий Git для разработки решений на основе языков программирования;
- применять язык SQL для прикладных решений;
- применять язык программирования Python и библиотеки при разработке решений на основе ИИ;
- использовать основные статистические методы анализа данных;
- визуализировать анализируемые данные;
- проводить EDA — разведочный анализ данных;
- применять методы машинного обучения для работы с большими данными;
- строить модель для прогнозирования временного ряда;
- строить модель для рекомендательных систем;
- использовать контейнеризацию и разворачивать модель в производство;

- применять технологии компьютерного зрения для обработки изображений и видео, задач биометрии;
- применять технологии обработки естественного языка для анализа и генерации текстов;
- использовать цифровые платформы анализа данных для решения профессиональных задач.

Владеть:

- языками программирования (Python) для реализации методов анализа и структурирования данных;
- основными инструментами для работы с данными (Git, Docker, RestAPI, Jupyter notebook, kaggle, Nvidia Cuda, VS, IntelliJIdea);
- навыками работы с основными библиотеками Python для визуализации данных (Matplotlib, Seaborn, Plotly);
- навыком создания логической и физической базы данных;
- навыком BI-аналитики;
- навыком работы с Numpy;
- способами улучшить качество модели с помощью методов Feature engineering;
- навыком использования библиотеки PyTorch, применения GPU в PyTorch;
- навыком построения полносвязной нейронной сети для задачи классификации;
- навыком работы с генеративными моделями GAN для задач генерации;
- навыком работы с автокодировщиками для задач уменьшения размерности и поиска аномалий;
- навыком работы с технологией RL — обучения с подкреплением;
- языками запросов для обращения к СУБД;
- основными принципами, возможностями и порядком применения методов обучения с учителем и без учителя для решения задач классификации данных;
- методами машинного обучения и нейронных сетей;
- математическими методами анализа данных;
- культурой постановки и планирования последовательности решения задач анализа данных и классификации;
- навыками анализа реальных задач из различных предметных областей на уровне отдельных подходов и коллективами алгоритмов;
- навыками работы с библиотеками, программными платформами (фреймворками) и программными комплексами машинного обучения.

7. СПИСОК ЛИТЕРАТУРЫ

Основная литература:

1. Бураков, М.В. Системы искусственного интеллекта: Учебное пособие / М.В. Бураков. – М.: Проспект, 2019. – 440 с.
2. Карпова, И.П. Базы данных: Учебное пособие / И.П. Карпова. – Санкт-Петербург: Питер, 2021. – 240 с.
3. Мартин Клеппман. Высоконагруженные приложения. Программирование, масштабирование, поддержка / М. Клеппман – СПб: Питер, 2018. – 740 с.
4. Кузнецов, В.А. Системный анализ, оптимизация и принятие решений: Учебник / В. А. Кузнецов, А. А. Черепяхин – Москва: КУРС, 2018. – 256 с.
5. Мышев, А.В. Архитектура для интеллектуальных вычислительных и информационно-измерительных систем с нечеткой средой вычислений // Федеральный институт промышленной собственности, отделение ВПТБ – 2019. – 29 с.
6. Алекс Петров. Распределенные данные. Алгоритмы работы современных систем хранения информации / А. Петров – СПб: Питер, 2021. – 336 с.
7. Дэви Силен, Арно Мейсман, Мохамед Али. Основы Data Science и Big Data. Python и наука о данных / Пер. с англ. Матвеев Е. – СПб.: Питер, 2017. – 336 с.: ил – (Серия «Библиотека программиста»). ISBN 978-5-496-02517-1.
8. Конвински Энди, Венделл Патрик, Захария Матей, Карау Холден. Изучаем Spark. Молниеносный анализ данных – Москва: ДМК Пресс, 2015. – 304 с.

Дополнительная литература:

1. Брокман Д. Что мы думаем о машинах, которые думают: Ведущие мировые учёные об искусственном интеллекте / Пер. с англ. М. Исаков – М.: Альпина Паблишер, 2017. – 550 с.: ил.
2. Бруссард М. Искусственный интеллект: Пределы возможного / Пер. с англ. Е. Арье – М.: Альпина Паблишер, 2020. – 362 с.
3. Бычков, А. Г. Сборник задач по теории вероятностей, математической статистике и методам оптимизации: учебное пособие / А.Г. Бычков. — Москва: Форум: ИНФРА-М, 2022. – 192 с. – (Среднее профессиональное образование).

4. Душкин, Р.В. Искусственный интеллект / Р.В. Душкин – М.: ДМК Пресс, 2022. – 380 с.: ил. – ISBN:978-5-97060-787-9
5. Фальк Ким. Рекомендательные системы на практике Практическое пособие / Пер. с англ. Д. М. Павлов. – Москва: ДМК Пресс, 2020. – 448 с.
6. О’Коннелл М. Искусственный интеллект и будущее человечества / Пер. с англ М. Кудряшова – М.: Бомбора, 2019. – 280 с.: ил.
7. Мартин Ф. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей / Пер. с англ. И. Рузмайкина. – СПб: Питер, 2019. – 400 с.: ил.
8. Нархид Ния, Шапира Гвен, Палино Тодд. Apache Kafka. Поточковая обработка и анализ данных – СПб: Питер, 2021. – 320 с.: ил.
9. Поляков, В. М. Методы оптимизации: учебное пособие / Поляков В. М., Агаларов З. С. — Москва: Дашков и К, 2022. — 86 с. — ISBN 978-5-394-05003-9.
10. Рязанов, С. И. Искусственный интеллект как множество — классификация искусственных интеллектов // Вузовская наука в современных условиях: сборник материалов 54-й Научно-технической конференции: в 3-х ч. Ульяновск: Ульяновский государственный технический университет, 2020. Ч. 1. С. 72-75.
11. Болотова, Ю.А. Методы и алгоритмы интеллектуальной обработки цифровых изображений: учебное пособие / Ю. А. Болотова, А. А. Друки, В. Г. Спицын; Национальный исследовательский Томский политехнический университет (ТПУ). — Томск: Изд-во ТПУ, 2016. — 207 с.: ил. — Библиогр.: с. 206. — ISBN 978-5-4387-0710-3.
12. Брэдшоу Шеннон, Брэзил Йон, Ходоров Кристина. MongoDB: полное руководство. Мощная и масштабируемая система управления базами данных. 3-я редакция / Пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2020. – 540 с.: ил.
13. Бруссард М. Искусственный интеллект: Пределы возможного / Пер. с англ. Е. Арье – М.: Альпина Паблишер, 2020. – 362 с.
14. Томас Дэвенпорт. Внедрение искусственного интеллекта в бизнес-практику. Преимущества и сложности / Пер. с англ. З. Мамедьяров. – М.: Альпина Паблишер, 2021. – 316 с.
15. Маркус Г. Искусственный интеллект: Перезагрузка. Как создать машинный разум, которому действительно можно доверять / Г. Маркус, Э. Дэвис. – М.: Альпина Паблишер, 2021. – 328 с.
16. ПикOVER К. Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей / Пер. с англ. А. Ефримова. – М.: Синдбад, 2022. – 250 с.

17. Цзэн Мин. Как Alibaba использует искусственный интеллект в бизнесе: Сетевое взаимодействие и анализ данных / Пер. с кит. В. Ионов. – М: Альпина Паблишер, 2022. – 360 с.: ил.

18. Сергеев, Л. И. Цифровая экономика: учебник для вузов / Л. И. Сергеев, Д. Л. Сергеев, А. Л. Юданова ; под редакцией Л. И. Сергеева. — 2-е изд., перераб. и доп. — М.: Издательство «Юрайт», 2023. — 437 с. — (Высшее образование). — ISBN 978-5-534-15797-0.

Учебно-методические и информационные материалы:

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

2. Большие данные в социальных и гуманитарных науках: Сб. обзоров и рефератов / РАН. ИНИОН. Центр науч.-информ. исслед. по науке, образованию и технологиям; отв. ред. – Гребенщикова Е. Г. – М., 2019. – 193 с. – (Сер.: Наука, образование и технологии). ISBN 978-5-248-00912-1.

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

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

5. Цифровая экономика от теории к практике: как российский бизнес использует искусственный интеллект [Электронный ресурс] / Исслед. РАЭК / НИУ ВШЭ при поддержке Microsoft. – 2019. – 66 с. – Режим доступа: <http://raec.ru/upload/files/190715-ii.pdf> (Дата обращения: 02.04.2024).

6. Бизнес переходит на искусственный интеллект / РБК+. Решения. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс] – Режим доступа: <https://plus.rbc.ru/news/638ce98f7a8aa9f3126daaa2> (Дата обращения: 02.04.2024).

7. Расставить нейросети / РБК+. Инновации. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс] – Режим доступа: <https://plus.rbc.ru/news/638ce67f7a8aa9e27b22f26e> (Дата обращения: 02.04.2024).

8. Технологии позволяют учитывать специфику каждой отрасли / РБК+. От первого лица. #1 Искусственный интеллект, 5 декабря 2022. – [Электронный ресурс] – Режим доступа: <https://plus.rbc.ru/news/638ced6c7a8aa9e28f7bf148> (Дата обращения: 02.04.2024).

Статьи:

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

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

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

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

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

6. Коженков, А.О.; Середа, А.М.; Поливаев, О.И. Искусственный интеллект в современном автомобилестроении / В сборнике: Теория и практика инновационных технологий в АПК. Материалы национальной научно-практической конференции. – Воронеж, 2022. – С. 210–216.

7. Красов, А.В.; Штеренберг, С.И.; Фахрутдинов, Р.М.; Рыжаков, Д.В.; Пестов, И.Е. Анализ информационной безопасности предприятия на основе сбора данных пользователей с открытых ресурсов и мониторинга информационных ресурсов с использованием машинного обучения [Электронный ресурс] // T-Comm. 2018. №10. – Режим доступа: <https://cyberleninka.ru/article/n/analiz-informatsionnoy-bezopasnosti-predpriyatiya-na-osnove-sbora-dannyh-polzovateley-s-otkrytyh-resursov-i-monitoringa> (Дата обращения: 02.04.2024).

8. Куницкая, О.М. Правовое регулирование искусственного интеллекта применительно к задачам развития экономики / В сборнике: Право и инновации: новые вызовы технологической революции. Материалы II Приволжского юридического конгресса. – Уфа, 2022. – С. 77–91.
9. Лазарева, М.М.; Калюкарин, А.В. Преимущества искусственного интеллекта. Применение в банковской сфере // Вопросы устойчивого развития общества. – 2022. – № 9. – С. 312–316.
10. Лазько, Н.В. Применение искусственного интеллекта в нефтегазовой отрасли // Мехатроника, автоматика и робототехника. – 2023. – № 11. – С. 155–158.
11. Львович, И.Я. Проблемы методологии проектирования интеллектуальных информационных систем // Информационные технологии в управлении, автоматизации и мехатронике. – 2020. – С. 120–123.
12. Машошин, А.И. Применение искусственного интеллекта при создании систем управления силами ВМФ // Морская радиоэлектроника – 2022. – № 2 (80). – С. 22–25.
13. Нарушения при работе с информационными системами Россельхознадзора будет выявлять искусственный интеллект // Пищевая промышленность. – 2022. – № 3. – С. 76.
14. Пиляй, И.В. Применение систем искусственного интеллекта для оценки времени и стоимости строительного проекта // Строительство и архитектура. – 2023. – Т. 11. – № 1. – С. 19.
15. Репин, И. С. Перспективные направления использования искусственного интеллекта в оперативно-розыскной деятельности // Журнал правовых и экономических исследований. – 2022. – № 2. – С. 125–130.
16. Субботин, А.В.; 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.
17. Шиллер, А.В. Место этической системы в архитектуре искусственного интеллекта // Вестник Томского государственного университета. – 2020. – №. 456. – С. 99–103.

Электронные образовательные ресурсы

1. Воронцов, К. В. Машинное обучение: курс лекций // MachineLearning.ru. – [Электронный ресурс] – Режим доступа:

[http://www.recognition.su/wiki/index.php?title=Машинное_обучение_\(курс_лекций%2С_К.В.В_оронцов\)](http://www.recognition.su/wiki/index.php?title=Машинное_обучение_(курс_лекций%2С_К.В.В_оронцов)) (Дата обращения: 02.04.2024).

2. Проектирование баз данных: распределенные базы и хранилища данных. Агрегирование // Национальный открытый университет «ИНТУИТ». – [Электронный ресурс] – Режим доступа: http://www.intuit.ru/studies/professional_retraining/953/courses/214/lecture/5508/ (Дата обращения: 02.04.2024).

3. Бесплатные материалы по Data Engineering от преподавателей МФТИ // МФТИ – [Электронный ресурс] – Режим доступа: <https://fpmi-edu.ru/free-de> (Дата обращения: 02.04.2024).

4. Курс «Машинное обучение» // Coursera. – [Электронный ресурс] – Режим доступа: <https://ru.coursera.org/specializations/gcp-data-machine-learning#courses> (Дата обращения: 02.04.2024).

5. Курс Big Data Specialization // Coursera. – [Электронный ресурс] – Режим доступа: <https://www.coursera.org/specializations/big-data> (Дата обращения: 02.04.2024).

6. Курс «Big Data и Data Science: начни погружение с нуля» // Stepik. – [Электронный ресурс] – Режим доступа: <https://stepik.org/course/101687/promo> (Дата обращения: 02.04.2024).

7. Open Machine Learning Course // mlcourse.ai. – [Электронный ресурс] – Режим доступа: <https://mlcourse.ai/book/index.html> (Дата обращения: 02.04.2024).

8. NLP-курс от ШАД // Lena Voita. – [Электронный ресурс] – Режим доступа: https://lena-voita.github.io/nlp_course.html (Дата обращения: 02.04.2024).

9. Бесплатный курс по Deep Learning от МФТИ // ФПМИ МФТИ. – [Электронный ресурс] – Режим доступа: <https://dls.samcs.ru> (Дата обращения: 02.04.2024).

10. Обучение Python и его библиотекам // W3Schools Spaces – [Электронный ресурс] – Режим доступа: <https://www.w3schools.com/python/default.asp> (Дата обращения: 02.04.2024).

Электронные информационные ресурсы

1 Курсы от образовательной платформы ООО «1Т» – [Электронный ресурс] – Режим доступа: <https://sprint.1t.ru/> (Дата обращения: 02.04.2024).

2. Искусственный интеллект // РБК – [Электронный ресурс]. – Режим доступа: <https://trends.rbc.ru/trends/tag/ai> (Дата обращения: 02.04.2024)

3. Искусственный интеллект (ИИ). Artificial intelligence // (AI) «TAdviser». Государство. Бизнес. Технологии. – [Электронный ресурс]. – Режим доступа: [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%B%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%B%D0%B5%D0%BA%D1%82_(%D0%98%D0%98,_Artificial_intelligence,_AI)) (Дата обращения: 02.04.2024).